Curtin Business School
School of Information Systems &
School of Management and Marketing

Frameworks for Evaluating Affective Design and Managing Prosumer Community Groups Using Blockchain and ML

This thesis is presented for the Degree of PhD
at
Curtin University

March 2021

Declaration

To the best of my knowledge and belief this dissertation does not contain material published by any other person except where due acknowledgement has been made.

This thesis does not contains any material which has been accepted for the award of any other degree or diploma in any university.

Signed:

Date: 28/03/2021

Abstract

This dissertation presents two novel frameworks in two different areas. Chapter 2- 6 presents APCG framework and Chapter 7 presents SMAF framework. Additionally a roadmap to guide machine learning problems for health industry is developed and presented in chapter 8. The two frameworks are as follows

1. Increasing Greenhouse Gas (GHG) emission is a question for future environmental sustainability. Viewing agriculture and land use management as a sink to greenhouse gases, this dissertation introduces a novel concept of "agro prosumer community group framework" which is potentially an effective way for locals to participate in carbon reduction mechanism while trading carbon-tokens and providing food traceability platform for local market. The framework is divided into four sub-frameworks to provide a smooth structure for APCGs. One sub-framework is to conceptualize APCG, the second one is an engagement framework for new prosumers, a third manages APCG and a fourth assesses the total amount of carbon captured by the APCG for trade with companies. Thus, a novel agro-prosumer network is proposed which provides a seamless structure and enables local participation in carbon reduction processes.
2. Rapidly changing world has increased the demand and requirement for a new product. A new product is now required to satisfy users' emotional desires along with elementary functions. Affective design plays an important role in the success of new products. Determining affective needs assist in understanding consumers' requirement. Since affective needs are subjective and varies individually, evaluating affective needs contains certain amount of uncertainty or fuzziness. Second framework propose a social media analytical framework, namely SMAF that utilizes consumer opinion data from social media in order to measure affective qualities in new product. The SMAF is integrated with fuzzy regression to predict affective quality; fuzzy regression is used to evaluate uncertainty which exists in users' affective quality perceptions. The mechanism of the proposed framework is demonstrated by a case study on affective design of automobiles.

Acknowledgement

I would like to express my gratefulness and special thanks to my Supervisor Dr Vidyasagar Potdar for his excellent supervision and constant support for this research. His suggestions, positive approach and excellent knowledge pool have been invaluable.

I would like to express my special thanks to my co-supervisor Dr Kit Yan Chan for his support, guidance and suggestions throughput the work done in this thesis.

I would like to say thanks to Dr Ponnie Clark and Dr Nik Thompson for their valuable reviews, insights for improving my research.

I would like to say thanks to my Internship mentors/ providers for giving me the opportunity to work on their projects and allow me to share my knowledge and briefly showcase the work done on the projects.

I am grateful to my peers and my school of Information Systems & School of Management and marketing for making it a great learning experience and smooth journey.

I dedicate my thesis to my parents Shashi & Suhagmal, my husband, Abhishek, and my two beautiful girls, Nayonika & Mishika for their unconditional support, love, inspiration and understanding. Without my husband's motivation and support, I would not have been able to handle challenges during this journey. I am very fortunate to have family by my side and living a happy life.

At last, I am thankful for my In-laws Shakuntala & Devendra, friends & family for their support, love and blessings.

List of Publications

Refereed Journal and Conference Publication [2017-2021]

[1] Conference paper presented at The International Conference on System Science and Engineering 2020 Pratima Jain and Kit Yan Chan , “A Social Media Analytical Framework Incorporating Fuzzy Regression for Affective Design”, Proceedings of The International Conference on System Science and Engineering, 2020

[2] Presented e-workshop abstract title- “An Agro-Prosumer Community Agricultural Framework for Carbon Sequestration and Trade using Blockchain Technology” at European Conference of Information Systems 2020 (ecis2020) for Blockchain workshop

[3] Presented conference paper titled- "Social Media Framework incorporating Affective Design: state of the art, challenges and opportunities" at 'The 2nd International Conference of Multidisciplinary Approaches on UN sustainable goals 2017, Thailand'.

[4] K.Y. Chan, C.K. Kwong, P. Clark, H. Jiang, Chris K.Y. Fung, B.A. Salih, Z.X. Liu, T.C. Wong, P. Jain, "Affective Design using machine learning: A survey and its prospect of conjoining big data, International Journal of Computer Integrated Manufacturing", 2018 (Accepted in August 2018). (ERA-B, 2016 impact factor: 1.459)

Contents

List of Tables & Algorithm

List of Figures

Chapter 1

Introduction

This chapter provides

1. The background of the research work
2. Background information on existing framework for harvest and carbon sharing community network
3. A summary of an agro-prosumer community group and use of Blockchain
4. The background of social media analytics and affective design
5. An outline of social media analytical framework and fuzzy regression techniques
6. The significance and importance of agro-prosumer community group and social media analytics for affective design
7. The research study's motivation and objective
8. The structure of this thesis
9. Main contributions of this thesis

1.1 Background of Research Work

The two major areas of research which are the focus of this study are the Agriculture-Prosumer community sharing network (APCG) and Affective design evaluation (SMAF). Additionally, two internship projects in health care industry were undertaken during the course of the PhD. The scope of the internship projects and a roadmap developed for the health industry are discussed in one chapter of this thesis. The frameworks for the aforementioned areas are briefly outlined below.

1.1.1 An agro-prosumer community group (APCG) framework for carbon and produce sharing network using Blockchain technology

A novel framework termed "Agro-Prosumer Community Group" (APCG) has been proposed to build a carbon-token and produce-sharing network using trust and transparency features of Blockchain technology as a design guide. The framework is divided into four sub-frameworks to provide a smooth structure for APCGs. One sub-framework is to conceptualize APCG, the second one is an engagement framework for new prosumers, a third manages APCG and a fourth assesses the total amount of carbon captured by the APCG for trade with companies.

Thus, a novel agro-prosumer network is proposed which enables local participation in carbon reduction processes.

1.1.2 Evaluation of uncertainty in affective design using social media analytical framework (SMAF) and fuzzy regression techniques:

A unique framework called "Social Media Analytical Framework" (SMAF) is developed to extract data from social media platforms. Social media data related to the affective design of a product is extracted, and fuzzy regression techniques are applied to accurately predict uncertainty present in the evaluation. A case study is undertaken of a luxury family car to demonstrate SMAF and predict the amount of uncertainty present in the evaluation of the car's affective design.

1.1.3 Roadmap to assist prediction model development in health industry:

In addition, current research techniques and machine learning concepts were applied to big health data as part of the internship project undertaken for the health industry. The problems and approaches associated with the project are the subject of one thesis chapter which describes the development of a new roadmap that facilitates the development of a prediction tool in health-related projects. The roadmap is a guidance to health industry ML and AI problems, and attempts to address the requirements and current shortcomings of this sector.

In this chapter both projects combined background has been presented, sections 1.2 and 1.3 provide the background information and solution outline for project A, i.e., "An agro-prosumer community group (APCG) framework for carbon and produce sharing network using Blockchain technology". Section 1.4 provides relevant background information, and section 1.5 gives an outline of the solution for project B i.e., "Evaluation of uncertainty in affective design using social media analytical framework (SMAF) and fuzzy regression techniques". Finally, project C, "Roadmap to assist prediction model development in the health industry", is briefly discussed in section 1.6.

Sections 1.7, 1.8 and 1.9 present the research motivation, objectives and significance, respectively and the structure of the thesis is set out in section 1.10. Section 1.11 concludes the chapter with the major contributions made by this research.

1.2 Background

The sequestration and reduction of greenhouse gas (GHG) emissions is essential for environmental sustainability. Industries, transportation, and the generation and consumption of electricity are the major contributors of GHG emissions (https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions); whereas, agriculture and land use management are making efforts to decrease GHG thus, act as sink to greenhouse gases.

Therefore, it is incumbent upon industries to take measures to reduce and mitigate carbon emissions. Various industries are working in this direction either by investing in carbon sequestration systems or making use of agriculture to reduce the amount of carbon. While carbon sequestration systems require substantial capital investments or advanced technology [1] finding an effective low capital carbon capture process by way of agriculture can be viewed as a better way to address this problem.

Since the agricultural industry is facing unprecedented problems in terms of climate change and a huge increase in food demand, it has to find solutions to the problems. In some cases, this is being done by using advanced techniques or chemical fertilizers to boost food production to meet the expected demand of an increasing world population. However, as a solution to climate change, the agricultural industry should apply ecologically-sound or traditional farming practices to reduce or sequestrate the ever-increasing GHG emissions. Alternative Food Networks (AFNs) play an important role in solving this problem, where environmental sustainability practices and food quality can be carried out simultaneously. In addition, AFNs can help to meet a small percentage of the world's food demand, and address long supply chain challenges such as security, food quality etc. Therefore, to deal with the aforementioned issues, various concepts such as AFNs are emerging; community supported agriculture (CSA) groups, carbon farming, urban farming and backyard farming as shown in figure 1.1 and discussed below briefly.

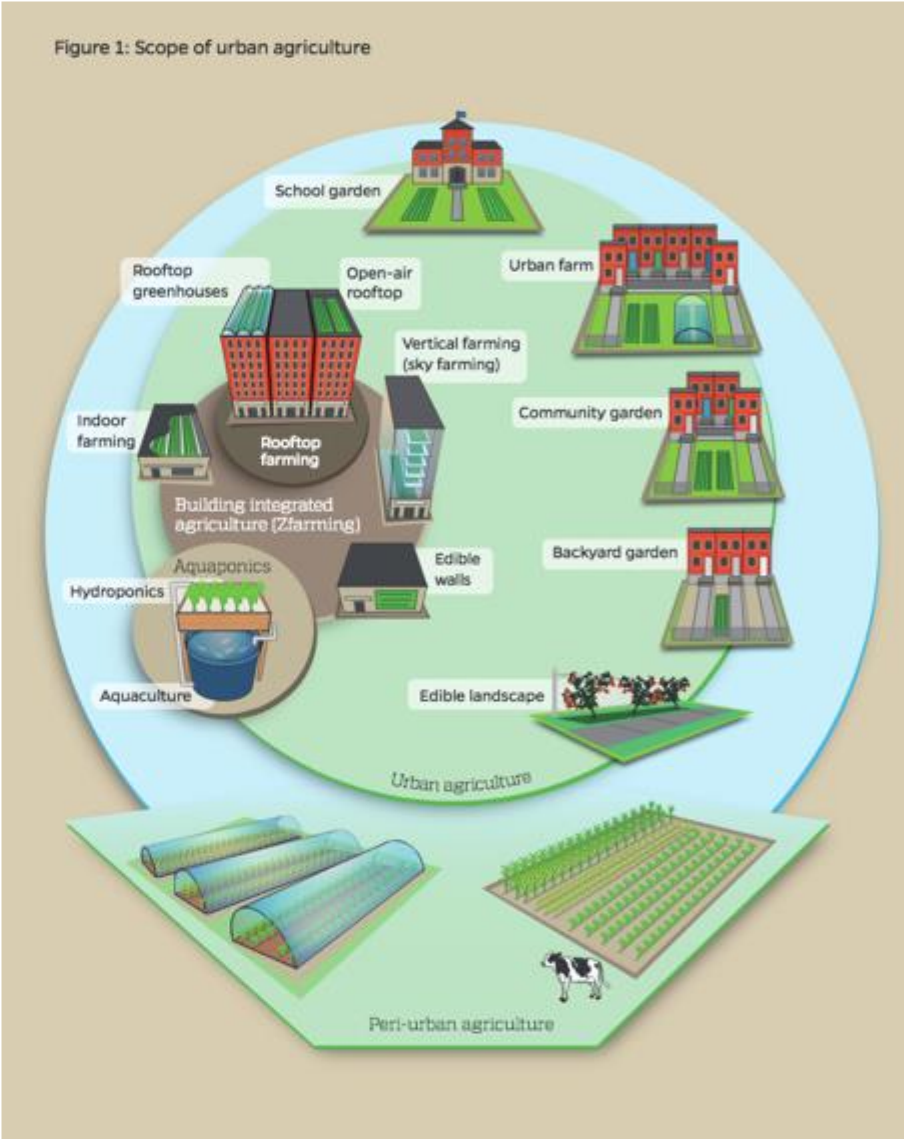

Figure 1.1 Alternate food networks

1.2.1 What is community agriculture?

Community agriculture is an AFN concept initiated in Japan by a small group of women who wanted quality milk and dairy products. The movement was started out of the rejection of unsatisfactory, poor quality milk and dairy products resulting from conventional agricultural practices [37]. The group then decided to support an alternative food supply to overcome inadequate industrial food produce. This movement drove the Community Supported Agriculture (CSA) concept which was later applied in Germany and Switzerland by small farmers who collaborated with local community members to create a member-farmer group. The CSA model has become quite popular in the USA, and now every state has at least one CSA farm.

"CSA is a partnership between farmers (or producers) and consumers (or members) to share benefits, financial risk and products of farm activities. The concept of CSA works in a manner, where farming cost is distributed among the members or stakeholders and final produces are consumed by members as shown in figure 1.2. In short CSA creates a direct link between farmer and consumers without government's aids or support" [39].

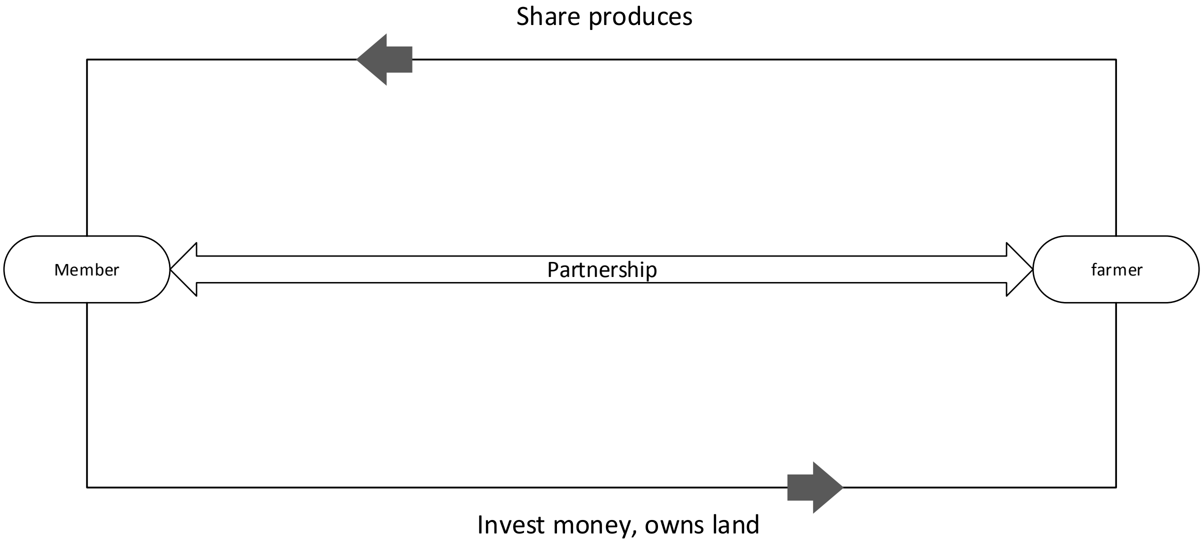

Figure 1.2 CSA model

In order to provide AFNs, several models of CSAs such as urban farming, vertical farming, Zfarming, community gardens and prosumption have been designed for a number of reasons. For instance, consumers want fresh and quality vegetables; they are now more concerned about food security; and more consumers are interested in eating food that is locally grown or produced. Furthermore, many people have started thinking of ways to improve environmental sustainability. Several CSA models are discussed below.

1.2.1.1 Urban farming/backyard farming/vertical farming

Urban farming is mainly driven by the passion to reconnect food production and consumption [40]. Urban faming or vertical farming is depicted usually as the utilisation of buildings, rooftop spaces and farming in city structures for food production and sustainability. Zfarming is another type urban of farming, where farming is done without using land.

1.2.1.2 Community gardens

"Community gardens" is the concept of creating places for members of the local community to grow plants, learn skills and meet other people. Community gardens are spreading worldwide primarily for food production, income, the creation of employment, and to improve local environments [41]. Community garden lands are usually owned by government, local council, school, church or housing agents.

1.2.1.3 Carbon farming

'Carbon farming' refers to land use and land management practices for carbon reduction and carbon sequestration. This is done by applying agricultural techniques whereby carbon is captured using natural sink sources such as soil and plants [7].

1.2.1.4 Prosumption

Prosumption is an old work culture where people used to grow their own vegetables and fruits. Prosumption is re-emerging in the current post-industrial era and becoming a concept where production and consumption meets. Food prosumers are individuals who actively produce food for self-consumption [25]. By producing food, prosumers act as both producers and consumers. They belong to a growing group of people who are interested not only in high convenience and low pricing, but also in the way that food is produced.

1.2.2 Why build a prosumer community group?

Toffler coined the term "prosumer" to describe a person who engages in producing food or goods for self-use. Although the industrialization era created a market where producers and consumers are two distinct entities, prosumption is making its way back through the demand for sustainability and a safe environment (https://www.scp-centre.org/promona/). "A prosumer is a producer and consumer in one person" [5]. In modern times, prosumption is a choice made by people who are concerned with food safety and a sustainable environment. Prosumers are working either for sustainability or to create an alternative food system at a better price and for better quality.

With the increasing interest in food quality, urban prosumers or food prosumers can be seen as a major source of high quality products and trustworthiness [3, 5, 20]. Prosumers are considered to be uniquely qualified minds that can unlock new value or trend in their markets [20]. In the domain of agriculture, prosumption is an economic, social and sustainable way to produce quality items at a low price by means of ecologically sound techniques. A prosumer community group will improve the economic value as well as build rich socio-psychological experiences (Jessen) among the prosumers by, for example, creating social relationships, acquiring self-pride, learning new skills, generating knowledge, and contributing to community work. Also, prosumers can obtain better value and exposure in market by forming a prosumer community group. Also it provides access to quality fruits/vegetables at better price thus resolving food insecurity for local community.

In addition to high quality produce, prosumers can generate carbon tokens indicating the total amount of carbon content absorption captured during the vegetation process, and trade it with the industries. However, in order to thrive as a group, the prosumer community has to perform a number of functions such as recording prosumers' information, predicting an individual's total amount of production, estimating an individual's consumption and, for the whole network, estimating the amount of carbon captured by individual prosumers to generate carbon tokens and keeping track of transaction details within the network or between the network and external buyers/companies. In view to these undertakings, it is vital to define and manage prosumer community groups. Thus, building and managing a viable agro-prosumer community group within an alternative food network is a promising step taken towards sustainability.

1.2.3 Managing Prosumer Community using Blockchain

In a prosumer community group, the prosumer is a key stakeholder who is encouraged to produce vegetables and generate carbon tokens. Prosumption is encouraged by incentives such as income earned from selling surplus carbon and vegetables, or social concerns such as climate change and sustainability. Generally, prosumers work both individually to produce and sell the surplus to buyers, or in a group whose members produce individually but collectively share their produce in the market. However, working individually makes it difficult to keep oneself motivated, and receive good exposure and value in market. The main reason for this is the unreliable nature of individual prosumers' output, as this can vary from season to season and often depends on environmental factors. Therefore, the individual prosumer often struggles to meet demand, making it impossible to offer bargain prices. Thus, they either have to settle for a low price or opt out of the market. Also, for carbon sequestration, one prosumer alone cannot earn enough carbon credits. Hence, the key motivation for a prosumer community group is that the prosumers can produce vegetables individually, but collectively share the combined amount in the market for better profit. Similarly, carbon tokens can be collectively shared with industries to obtain substantial carbon credits.

Currently, farmers belonging to a community group utilise a large tract of land to form a group; on the other hand, urban prosumers undertake to grow edible plants for themselves [37, 39]. However, it is difficult to form a community model comprised of prosumers who own backyard or small garden spaces, and current attempts are still in the early stage. Because of the desire for safe environment practices, locally grown food and food safety is a growing interest. Hence, there is the need for a model that is appropriate for agro-prosumers or backyard farmers in community groups – a model that is lacking in current community agriculture groups. Additionally, carbon-reduction farming practices along with the benefits of community agriculture can be combined to take advantage of the sequestration of carbon in farming and the good quality produce obtained through CSAs. Thus, the formation of a community of urban agro-prosumers can promote them as a group, give them good exposure and the motivation to apply environmentally friendly and sustainable farming practices.

The efficient management of a prosumer community group can produce several benefits. For instance, it can help to ensure that prosumers remain goal oriented, encourages them to earning good incentivising rates or income, helps them to source external buyers or stores/organisations that will buy their produce and/or carbon tokens, and motivates them to apply ecological practices for carbon capture while recruiting new prosumers for the group, thereby helping to create to a viable network.

A group or network generally has a central body that performs management-related tasks and transactions. However, a decentralised platform is in demand due to various issues that have arisen in central management systems such as insecurity in current agriculture system, lack of tracking to ascertain food origins, lack of fresh, high quality produce, and increase in fraud or double-dealing during the goods sharing process. Thus, the notion of forming a prosumer network via decentralized platform is an attractive prospect. Therefore, one aim of this study is to examine the concept of a prosumer community group in the agriculture domain, where the focus is on designing a framework that will facilitate the establishment of a sustainable network and engage prosumers in a number of activities that will keep them motivated and goal-oriented. In this research undertaking, a novel framework is presented for the establishment of an Agro-Prosumer Community Group (APCG) which will enable prosumers to develop socio-economic relationships while contributing to improved food safety, efficient food traceability, and environmental sustainability.

1.3 Outline: Agro-Prosumer Community Group (APCG)

An Agro-Prosumer Community Group (APCG) is a network sharing group, whose members have similar agricultural production behaviours and goals, compete in the market as a group as shown in figure 1.3. The main reason for establishing an APCG is to provide a sustainable, trustworthy and local community group, where socio-economic relationships and technology provide a platform for earning incentives while adhering to carbon sequestration principles.

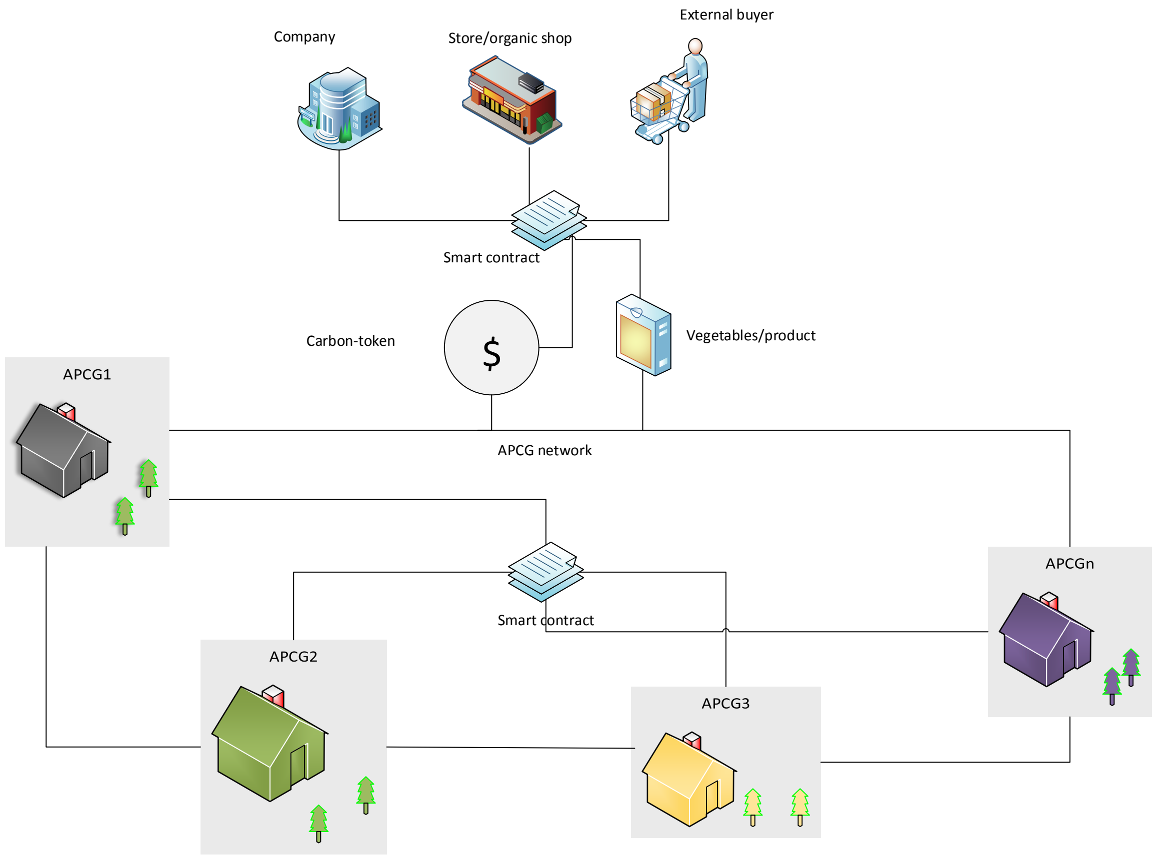

Figure 1.3 Agro-prosumer produce sharing process

APCGs provide a means to interact with each other and with the local council, retailers, supermarkets, companies/organizations, and other consumers via web applications. However, it is notable that the current literature review shows no evidence of prosumer community groups in the agricultural domain being considered together with a token mechanism for trading carbon with industries. This study has been motivated by this research gap. Section 1.4 provides background information on affective design evaluation background. This is followed by an outline of the solution in section 1.5.

EVALUATION OF UNCERTAINTY IN AFFECTIVE DESIGN USING SOCIAL MEDIA ANALYTICAL FRAMEWORK (SMAF) AND FUZZY REGRESSION TECHNIQUES:

1.4 Background

In new product development, other than the product's basic functionality, the factor which is attracting more interest is the item's affective or aesthetic qualities. For instance, many users nowadays are more concerned with the colour, shape, look and feel of a product (i.e., the affective qualities of cars, smart phones etc.). Hence, affective design plays an essential role in new product development.

The consumer's voice plays a critical role in the development or evaluation of a new product's affective design. The popular social media channels such as Twitter, Facebook and Instagram are a significant source of information and public content, and provide platforms for information exchange [42]. Social media data consists of users' opinions, comments, emotions and influences, contributing to a massive amount of knowledge-sharing in various domains such as tourism, government, politics and product co-creation [43]. Social data contains useful information for identifying affective preferences regarding new products [44]. From social media data, it is possible to extract informative opinions regarding the affective satisfaction derived from a product. Thus, using social media data analytics, the voices of consumers can be extracted from contents related to affective design, which are shared on social media [45].

Social media data is essentially a collection of consumer-generated data found in blogs, posts, comments, reviews and other forms of communication which are created by consumers. Since social data is influenced by personal feelings, tastes, and opinions [44], it is inherently ambiguous and uncertain. In today's highly competitive market, it is essential to evaluate users' perceptions of affective design along with the amount of inherent uncertainty. For instance, a new product is not perceived as being just 'good' or 'bad'. The users' judgement and opinion of affective design can be expressed as 'I like it', 'I love it', 'I am not so sure', 'LOL' or even by emoticons such as a thumbs up, smile, grumpy expression etc. Thus, it is important to evaluate the uncertainty present in social data.

1.4.1 Social Media Data

These days, it is difficult to imagine a world without the Internet of which social media usage is an after-effect. Expansion of mobile technologies, 3G/4G network and Internet has been serving as the primary force to provide technological platforms for information dissemination, content generation, and interactive communications [42]. Social media usage is one of the most popular online activities and is drawing enormous attention from both application and research perspectives. Social platforms not only serves as an integral part of information ecosystems, but its usage by individuals, consumers, corporations, governments and many other entities, is increasing rapidly [42]. This trend has revolutionised the power of expression in recent times. Social media is currently being utilised worldwide at an exponential rate. Individuals spend more and more time now on these sites and it is constantly increasing. For example, in 2020, over 3.6 billion people worldwide were using social media, which is projected to increase to almost 4.41 billion by 2025 (https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/). Facebook, YouTube, Twitter, Instagram and LinkedIn are just a few of the most popular social media web applications. The amount of time people spend on these sites is alarming now. For instance, Australians spend an average of one hour 47 minutes per day on social media, and this has been increasing slowly over the past eight years (https://www.genroe.com/blog/social-media-statistics-australia/13492). The enormous growth of social media usage has led to the generation of massive amounts of data, which has been collectively known as 'social media big data' or 'social big data'. Hence, social media landscapes can be viewed as widespread communication platforms for the consumption and sharing of an undertaking, a rich source of day-to-day information.

Social media plays an increasingly significant role in an individual's social life by introducing enhanced features related to their emotions and behaviour. For example, Facebook allows users to update their status or post information not only in the form of text, but also as graphic emotion icons. Thus, different forms of data that includes textual data, pictures, videos, sounds, and geolocation are provided by social media. By way of illustration – suppose that an individual driving a newly-released car model had recently updated his views about it on a social media platform. His feelings and behaviours regarding the car can now be analysed based on his comments and the use of any emotion icons. Generally, the social media data can be divided into unstructured data and structured data, where textual content and graphic content is unstructured and followers/friends are considered as structured data [46]. Collectively, this information constitutes online knowledge and can be analysed to obtain insights related to products, services and brands, that are shared by a range of different users [43]. Rathore, Ilavarasan, and Dwivedi state that social media is a major influencer of users' behaviour as it provides opinions, perceptions, feedback, usage, intention, purchase habits, depth of analysis and a variety of shared information. Hence, collecting information from these interrelated multiple sources opens opportunities for acquiring more reliable and accurate knowledge which can further contribute to better business insights [44, 47]. With social networks, designers can use online consumer data to understand their requirements, affections and desires. Hence, extracting and incorporating emotional, behavioural and demographical social media texts into the product design process may contribute to enhancing the affective design features.

1.4.2 Affective Design

Affective design involves the processes of identifying, measuring, analysing and understanding the relationship between the affective needs of the customer and the perceptual design attributes in the design domain [48]. The purpose of affective design is to satisfy the affective needs of users, by integrating their affective requisites in the design attributes of a new product [49]. This enables designers to create product designs which appeal more to their target market. Affective design has been shown to stimulate customers' psychological feelings and can help improve customer satisfaction [50]. Conventional methods of affective design focus on predicting or determining the amount of affective quality. These methods include Neural Networks [51] and the Kansei Methodology [52] which predicts consumers' perceptions when the perceptual design elements of a product are given. Kansei Affective Engineering ('KE') [53] is defined as the technology of incorporating the consumer's Kansei in the product design. According to Nagamachi, there are four steps in the Kansei Affective Engineering process: (1) grasp the consumer's Kansei in the specific product domain using psychological or psychophysiological measurements; (2) analyse the Kansei data by statistical, medical, or engineering methods in order to clarify the Kansei structure; (3) interpret the analysed data and transfer the data to the new product domain; and (4) design a new Kansei product. The KE methodology has contributed to the successful development of many different products such as motor vehicles, coffee cans, beer cans, milk cartons and body cosmetics [48, 52, 54, 55]. However, perceptions are subjective, which means that they contain a certain amount of uncertainty. For example, blogs, opinions or comments about products are never just good or bad, but contain terms such as "not sure", "I love it", "Lol", "I like it" etc. Thus, uncertainties are inevitable in human perceptions. Hence, there is the need to evaluate uncertainty present in the human perception of design. However, in the literature, little attention has been paid to the element of uncertainty in affective quality. Major works in this area include Chan et al. [49] who proposed an intelligent fuzzy regression technique to generate models for relating design variables to affective responses in which both non-linearity and fuzziness are considered. Chan and Engelke [56] presented a novel fuzzy regression method to predict affective quality and fuzziness in human assessments, for given objective features. However, no previous studies have utilised social media data (particularly social affective design data) to predict its inherent uncertainty using fuzzy techniques.

1.5 Solution outline: SMAF

The aim of this study is to utilise social big data to extract affective design opinions from users and predict the uncertainty present in the posted opinions for better evaluation. To evaluate the amount of uncertainty in human perceptions or subjective judgement of social media data, the proposed framework will incorporate fuzzy regression techniques.

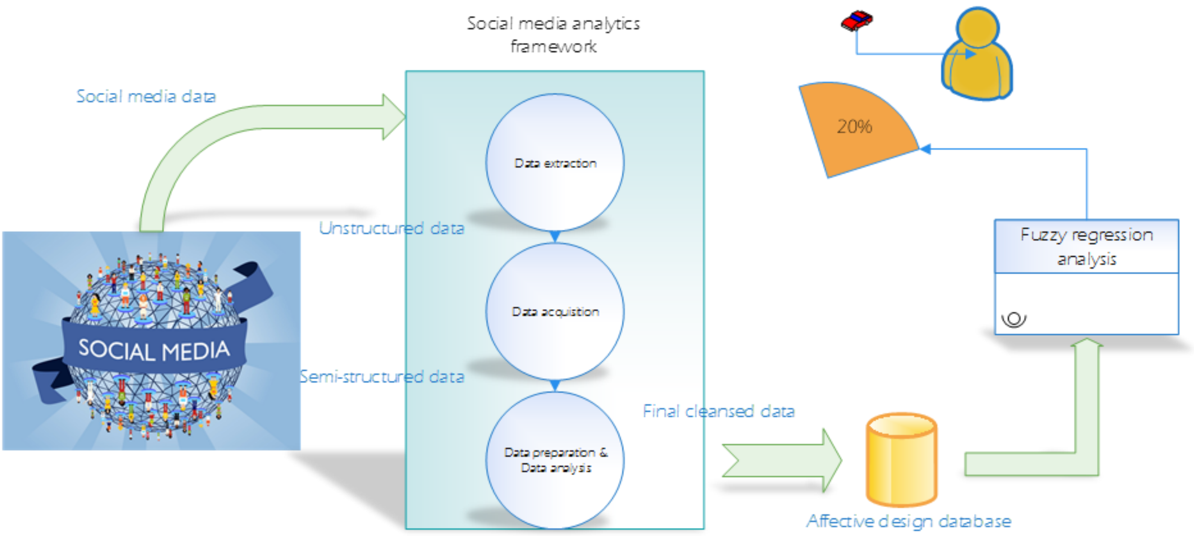

Figure 1.4 Social Media Analytical Framework (SMAF) for affective design evaluation

The proposed Social Media Analytical Framework (SMAF) illustrated in Figure 1.4 consists of two main components: social media data analytics and fuzzy regression analysis. The social media data analytics component is developed using the following phases of the big data cycle: data extraction, data acquisition, data storage and data preparation, and data analysis. These phases are intended to capture customer opinions from social media platforms. SMAF also extracts from the customer opinions the design attributes and affective quality. Because users' design perception contains fuzziness, it is important to determine the design perceptions accurately. Fuzzy regression techniques are utilised to predict uncertainty present in affective design perceptions, which is second component in this framework. Fuzzy regression analysis is used to analyse the magnitude of affective quality and uncertainty when the consumer evaluates the affective quality of a product.

The method is demonstrated with a case study involving the development of an automobile design. This study utilized a luxury family car as an automobile product and examined the affective design perceptions along with the evaluation of their inherent uncertainty for accurate evaluation. The colour feature of the car models is used as a design attribute, and results are discussed in detail.

ROADMAP TO ASSIST PREDICTION MODEL DEVELOPMENT IN HEALTH INDUSTRY

1.6 Background

In recent years, the development of computer software and hardware technologies has enabled the industry to capture massive amounts of data, otherwise known as ‘big data’. In addition, with the adoption of electronic medical systems, health-related data such as Electronic Health Record (EMR/EHR), medical claims and images have become readily available, which generates big data in health industry [57]. This “big health data” has huge potential to assist in healthcare decisions or treatments [57-60]. However, it is now evident that big health data is useless unless it’s potential is inferred, analysed or interpreted correctly to identify useful patterns or knowledge [61]. When applying ML and AI techniques to big health data, algorithms that could help with medical treatment or diagnosis, management care can be determined from datasets and then transformed into tools or acted upon as a service.

Two health industry projects were undertaken as part of the doctoral internship. The internship is done to apply research skills and machine learning concepts on health industrial projects. The internship project problems and approaches has been presented in the form of case studies. A new roadmap is developed as part of the thesis contribution to assist in developing prediction tool for health industry. The roadmap is presented as a guideline solution to health industry ML and AI problems, and is based on the requirements and limitations of this sector.

1.7 Research motivation

1.7.1 APCG Project:

Various research studies have been conducted on community agriculture and alternative food networks; however, the establishment of prosumer community groups has received little if any attention. The initial literature review conducted for this research revealed that there is no framework for a prosumer community agricultural group, although the literature contains community agriculture concepts such as urban farming, kitchen garden, are community gardens [10, 11, 15]. Additionally, the current alternative food network or CSA models do not include motivations for utilizing carbon reduction methods. Community agriculture groups are generally formed by small-scale farmers and community members [37] to meet local food demand and address issues of quality, however a network of local prosumers that can contribute towards carbon reduction mechanism while satisfying local food demand is a motivation of research here, following key gaps exist in current community groups are briefly explained below-

1.7.1.1 Involvement Deficiency of local prosumers in current community agriculture groups

Current community agriculture relies heavily on CSA members' participation and their high levels of satisfaction for its success. These groups include large-scale farmers or small-scale farmers who cultivate crops/vegetables, community groups have low levels of engagement by prosumers or urban gardeners. In fact, the previous studies show community groups for farmers or shows concept of individual prosumers/farmers. Also as discussed earlier, individual prosumers are not able to produce reliable amounts of food which would give them better market exposure. To address this, the formation of a community group for prosumers seems to be a positive prospect.

1.7.1.2 Dearth of management of community group and its structure

The concept of an agro-prosumer community group is still either unknown or immature; therefore, the existing literature has not much to offer in the way of literature reviews of associated concepts or resolutions for the linked challenges. Thus, prosumer community groups are not adequately represented in the literature. Hence a structure is required that facilitates the smooth operation of prosumer community groups and meets the key management needs of these groups. These issues are discussed below.

a) Definition and formation of agro-prosumer community group (APCG): To form a prosumer community group, agro-prosumer's production behaviour must be analysed to group them according to their production characteristics.
b) Lack of goal-oriented nature of community group: Current community groups are not goal-oriented or intrinsically motivated, which may lead to an inconsistent production. Thus, a goal-oriented management system must be developed for APCG.
c) Approaches for assessing new prosumers' behaviour and recruiting them for the APCGs: To create a sustainable group, new prosumers should become involved. Assessing new prosumers' real time production behaviour can help them to achieve a better and long-term relationship with an APCG. Thus, it is important to have a framework for new prosumers engagement based on an analysis of their production capacity.

d) Estimating carbon captured by prosumers and digital tokens for trading purpose: In order to motivate APCGs to embrace environmental sustainability, a carbon capture mechanism should be introduced to encourage their participation by, for example, by providing incentives for engaging in carbon trading and simultaneously helping external organisations to reduce their carbon emissions.
e) Lack of a transparent and traceable platform for network sharing: Current community group structures either have a central body or are dependent on community members' satisfaction for their success. A decentralised platform can resolve this problem and can also add a food traceability feature.

1.7.1.3 Absence of sustainable structure for current community agriculture practices

Since farmers are often not motivated to practise ecological farming because the yield is lower, they use techniques such as the application of chemical fertilisers or modern equipment to increase production. Although government is encouraging the agriculture sector to reduce carbon emissions and generate carbon credits to earn incentives, most farmers are not aware of the policy or its benefits. Thus, it is evident that there is little motivation for farmers to engage in ecological farming practices and goal-sharing objectives combined with carbon farming in community agriculture.

1.7.1.4 Limited ways to build trust and transparency in the framework

Currently there are certain issues in the community agriculture groups such as the lack of trust, traceability and transparency in the system. To solve trust and traceability issues, a decentralised system is required to implement the APCGs and a carbon framework. A decentralised technology such as blockchain has yet to be used for the purpose of creating a community agriculture structure for better markets, connections and income for urban farmers and prosumers. Thus, a formal structure design or an innovative model is required to establish and control the activities of community agriculture groups.

1.7.2 SMAF Project

The comprehensive study of the literature in the areas of social media analytics, data fusion, affective design and fuzzy regression reveals that in the current dynamic world of the Internet boom, manufacturing companies need to be aware of consumers' affective needs and perceptions when marketing or designing their products. These needs can be effectively captured from social media channels. Although a lot of work has been done in assessing affective design attributes, affective quality and the most preferred design values, little has been done to determine perceptual uncertainty in social media data related to affective quality evaluation. The literature review identified several concerns, which are discussed below.

1.7.2.1 Lack of utilizing dynamic social media data in product design while evaluating users' affective design perception/opinion

Current product design companies evaluate users' design perception and design needs by using survey techniques or building prediction models. Only a limited number of users are available to participate in a survey, or only a few features are published via a survey. Although a prediction model works well with a limited number of features, with given the increasing competition and the demands of modern consumers, companies must take advantage of dynamic social media data in order to analyse a variety of users' affective design needs and evaluate them accurately.

1.8 Research objectives

For APCG: The aim of this particular research project is to build an agro-prosumer community group (APCG) to address environmental sustainability issues and provide a local food-network-sharing group that can work as an alternative food network. To build the framework, an extensive literature review of all related research was undertaken, along with development solution to specific research issues and verification of the models. Also, blockchain technology is required as a design guide of the framework.

FOR SMAF: The aim of this research work is to build an analytical framework to extract, prepare and analyse the social media data related to affective product design. The affective database is then evaluated for amount of inherent uncertainty, utilizing fuzzy techniques.

The main objectives of this study are listed below.

1.8.1 To create a framework for to define APCG concept

To build APCGs, the first critical step is to conceptualize APCG. Because an APCG is a fairly a new concept in the domain of agriculture or alternative food networks, a detailed structure has not been proposed by previous studies. Therefore, one aim of this research is to develop a framework to define APCG concept. In this framework, prosumers are analysed and classified into different groups based on their unique characteristics.

1.8.2 To create an engagement framework for new agro-prosumers

In order to ensure the longevity of APCGs, it is important that as many prosumers as possible join their local APCG. Hence, it is important to recruit new prosumers for existing groups. Moreover, current studies show the need for a framework to encourage new prosumers to join agriculture groups. Thus, this study proposes a new framework for recruiting new prosumers for the group even if they have no farming knowledge or production history.

1.8.3 Make APCGs a goal-oriented network:

For this research objective, the aim is to establish a framework to manage goals and encourage APCGs to be a goal-oriented network. Thus, an effective framework is required to define diverse goals for APCGs. Goals can be contradictory, for example, one goal can be achieved at the expense of other goals. In this situation the goals can be negotiated in order to that each can be achieved satisfactorily. Hence, an ideal set of overall goals for APCGs is what is expected from a goal-oriented framework. A linear goal-programming model is used to address the problem of conflicting goals and establish an ideal set of goals.

1.8.4 To develop carbon token framework using Blockchain technology:

Previous researchers have proposed frameworks that address trust and transparency issues in agriculture by using blockchain technology. However, APCG is a new model for a prosumer community network that involves the novel concept of carbon token estimation and generation performed between APCGs and external entities. Therefore, the aim is to evaluate prosumers' carbon capture and implement a traceable platform for secure transactions and trading purposes using blockchain technology.

1.8.5 To develop a model using social media data and fuzzy regression techniques to evaluate affective design:

There is a need to predict affective quality and estimate human uncertainty when evaluating the affective design of new products as a means of accurately determining customer needs and demands. To achieve this objective, an innovative model is built to extract users' perceptions from social media and evaluate them using current fuzzy regression techniques.

Table 1.1 Pictorial presentation of research gaps and research objectives

S.No	Objective	Design Requirement	Proposed solution framework
1	To create a framework for to define APCG concept	How to define APCG concept?	Agro-Prosumer Community Group (APCG) concepts
2	To create an engagement framework for new agro-prosumers	How to engage new prosumers to the existing prosumer community groups?	Agro-prosumer Engagement Framework
3	Make APCGs a goal-oriented network:	How to motivate the prosumers to commit to the sustainable environment cause? How to maintain an economically self-sustained group?	Goal Management Framework
4	To develop carbon token framework using Blockchain technology:	How to maintain trust in prosumer networks, i.e., between prosumers and the market? How to ensure food traceability/carbon token transaction?	Carbon Assessment and Blockchain Framework for Carbon Sharing
5	To develop a model using social media data and fuzzy regression techniques to evaluate affective design:	There is a lack of research focusing on the estimation of uncertainty in affective quality evaluation using social big data and fuzzy regression techniques. Most research has focused on determining the magnitudes of affective qualities by consuming limited survey data.	A Social Media Analytical Framework Incorporating Fuzzy Regression for Affective Design presented with a case study a case study

1.9 Scope of the thesis

This thesis contains two frameworks. The first framework proposes the concept of an Agro-Prosumer Community Group where the structure of the APCG is present that provides framework to a number of problems or requirements such as (i) defining APCG concept, (ii) new agro-prosumer engagement process for APCG, (iii) APCGs goal management, and (iv) APCG carbon token evaluation using blockchain technology.

In the second framework, social media platforms are used to obtain affective product design data. The framework is termed Social Media Analytical Framework (SMAF), which includes data extraction from social media platforms, the transformation of data from unstructured to semi-structured format, and data cleansing which includes removing missing values, unwanted data and data analysis for design related information. The final prepared data is then used to evaluate uncertainty present in it by means of fuzzy regression techniques.

1.10 Significance of the research

It is anticipated that this project will have socio-economic and technical significance, explained below.

1.10.1 Social and economic significance

- Satisfies local food demand: With the growing demand for locally-produced food, APCGs provide a sustainable, social and economical way to connect prosumers within a local produce-sharing network to produce and share within or outside the network.
- Provides sustainable agriculture community groups: APCGs not only offer surplus vegetables and fruits to share within the network and with external buyers, but also promotes ecological ways of farming so that the network can share carbon tokens with organisations to offset their increased carbon emissions. Thus, APCGs provide an alternative sustainable agriculture network.
- Encourage local public participation in carbon sequestration: With APCGs, local community members can participate in carbon sequestration either through their own farming activities or by maintaining their backyards/garden spaces in ecological ways. APCG promotes carbon token generation, where carbon absorbed by trees can be used to offset prosumers' own carbon emission or support other organizations. Thus, the local population can also be a part of carbon sequestration system and contribute to their own country's carbon offset targets.
- Prosumers management and accessible system: APCG network is based on categorizing similar production behaviour prosumers. Therefore ease of income distribution or disagreements among members are solved to certain extent. Further, the external buyers or supermarkets can approach these local communities as a whole or as groups to make contracts with APCGs. This makes the process of accessing prosumers much more efficient, and for any quality issues, tracing and recalling would be much easier.
- Support prosumers to earn incentive while working on their interest: APCGs will increase employment by engaging local members as prosumers. Prosumers can earn in two ways: by selling produce or by sharing carbon tokens.

- Improve and build social and learning network: Prosumer networks can create a social platform where members can motivate each other and learn from each other, which helps to build a social and knowledge-sharing network.
- Utilise social big data for affective design evaluation: Popular social media platforms are being utilised to understand customers' need. This framework provides affective design perception contained in social media data, enabling companies to better design product features and predict customers' affective needs.

1.10.2 Technical significance

- Formulation of the concept of APCGs for produce and carbon sharing network: Agro-prosumer community groups (APCGs) are a new concept involving the creation of a produce- and carbon-sharing network, which can help to address two current problems: local food demand and carbon emission reduction. The few studies on community agriculture investigate mostly farmers, and the benefits of a member network or individual prosumers but, to date, no researchers have been concerned with the creation of prosumer community groups or the challenges they face.
- Presenting framework to define APCG concepts: The first step for APCGs is to define APCG concept by defining prosumers and categorize them according to their unique attributes. In this study, prosumer clusters are formed using the hierarchical clustering method (deep learning algorithm). Prosumers' previous production profiles are collected and analysed using this hierarchical method. This process identifies the characteristics of prosumers which helps to categorize them into groups, each of which can then be individually defined and labelled.
- Creating a new engagement solution for APCGs: In the next step, in order to introduce new prosumers to the APCGs, a framework is built which evaluates dynamic production behaviour to determine the selection of new prosumers. Here, prosumers can also choose a group to which they want to belong, then the prosumer is evaluated for a certain period and is accepted into the group if a specific production threshold is achieved.
- Proposes a framework for identifying overall APCG goals and managing conflicting ones to achieve goal-oriented nature in APCGs: APCGs have two main conflicting objectives: production of an adequate amount of vegetables and fruit, and carbon generation. In order to improve their yield, generally producers utilise modern

techniques or chemical fertilizers which can contribute to carbon emissions. However, ecological farming practices may result in less produce. This study presents an innovative framework comprising multiple criteria goal programming (MCGP) techniques for managing conflicting goals among prosumer groups. The use of MCGP approach has been used for prosumer community groups in smart grids, however it is never used in agriculture sector so far.

- Proposes a carbon token formation process and prototypes for APCGs: With the current concerns about increased global greenhouse gas emissions, APCGs present an innovative framework where local members can become prosumers and contribute to carbon sequestration or carbon emission reduction by capturing carbon in their backyard or garden trees/plants and share it collectively with organizations or government entities. For this study, a DERAF technique and carbon calculator is used to compute the amount of carbon absorption by trees which is then converted into carbon tokens to share between prosumers in the network and companies. This chapter also includes prototypes for APCG, particularly for the carbon token framework. This is the first framework to show how individuals can become involved in and contribute to environmental sustainability. Also, blockchain technology is used as a design guide for the framework.
- Manufacturing companies can enhance their product design process: By applying the proposed framework, companies can better understand users' affective preferences and specific design values for new product development by utilizing a wide range of dynamic consumer voices through a social media analytical framework.

- SMAF for data extraction; the social media analytical framework is a major contribution which can be utilized as a way of extracting keyword-based social media data and convert unstructured data into structured data for the purpose of data analysis.

1.11 Thesis structure

The structure of this thesis is organized as follows.

Chapter 2 presents a comprehensive literature review of APCGs and related research studies. In this chapter, we examine past research studies on community agriculture groups, backyard farming or urban farming, and prosumers involvement in it. Very few studies talk about prosumers and the creation of community groups for the purpose of exchanging produce and participating in a carbon-sharing network. Hence, this study addresses this gap with the concept of APCGs.

Chapter 3 presents the APCG concept that define agro-prosumers and classifies them on their production-sharing behaviours before allocating them to the appropriate APCG. The framework helps in determining the pre-requisites of different groups of APCGs. The framework is verified using prosumers dataset to seek the functionality of the model.

Chapter 4 presents a framework for engaging new prosumers in the APCGs. This framework is designed to assess new prosumers production performance by using an iterative evaluation process before assigning them the membership. Further, the framework is verified using prosumer dataset.

Chapter 5 proposes a framework to make APCGs a goal-oriented network. The framework identify and manage diverse goals of APCGs. Conflicting diverse goals are negotiated in order to achieve best solution for APCG i.e. most favourable set of goals. Further, a linear goal programming model is used to find the most favourable set of goals.

Chapter 6 proposes a framework to assess the carbon content captured by each prosumer and combined to share with the organizations to offset their carbon emissions. To achieve this, initially carbon calculations are done by using prosumers' information related to the number of trees and type of vegetation. The amount of captured carbon is converted into carbon tokens that can be traded with companies. Blockchain technology is used to design a framework which enables transactions between APCGs and external companies. Further, we design prototype to illustrate the functionality of the proposed framework using prosumers data set.

Chapter 7 presents a comprehensive literature review of social media posts and affective design. A social media analytical framework (SMAF) is proposed that utilizes consumer opinion data posted on social media in order to measure the affective qualities of new products. The SMAF is integrated with fuzzy regression to predict affective quality; fuzzy regression is used to evaluate the uncertainty inherent in users' affective quality perceptions. The mechanism of the proposed framework is demonstrated by a case study on the affective design of automobiles.

Chapter 8 describes machine learning implementation in detail and discusses some major aspects of ML such as model biases, methodological limitations, and interpretation of the models, in the context of the health care industry. A roadmap is developed which can be used as guide or enhanced when developing ML models in the health sector.

Chapter 9 concludes the thesis with a summary of what has been achieved and its major contributions to extant knowledge. It discusses the work that still needs to be done and offers suggestions for future research directions.

1.12 Main Contributions of this thesis

To date, previous researches have considered agriculture as a substantial domain for carbon sequestration. “Urban agriculture” or “vertical farming” are responses to the increasing demand for local grown vegetation, and potentially could provide carbon sequestration if designed properly [22]. Carbon farming is also seen as a valuable means of earning incentives for farmers [62]. However, the literature has little discussion on prosumer community agriculture groups as a sharing network. In this dissertation, a new concept of forming an agro-prosumer community group (APCG) has been proposed. Some of the current problems associated with agriculture are the lack of environmental sustainability, carbon emission reduction mechanism, traceable network sharing, and the availability of fresh local food. These problems can be solved by developing agro-prosumer community groups. This study presents four frameworks and solution model for each to build an APCG network using blockchain as a design guide.

Another major contribution of this thesis is the development of a unique framework which extracts data from social media platforms such as Twitter and Instagram. The proposed solution is a social media analytics framework (SMAF), that integrates fuzzy regression techniques, for predicting uncertainty in evaluating affective design. A case study is used to demonstrate the solution. Along the same lines, various aspects of ML are discussed, and machine learning approaches are utilised in the health industry to build prediction models using big data from the health domain. Finally, another key contribution of this study is the roadmap developed to assist the health industry to develop reliable prediction tools.

Chapter 2

2 Literature Review: An Agro-Prosumer Community Group for Carbon Capture and Trade using Blockchain Technology

This chapter provides

- A review of the literature pertaining to alternative food networks and prosumer involvement in agricultural community groups.
- A detailed study of existing approaches and their adaptability to manage and form a prosumer community group for carbon sequestration and a produce-sharing network.
- An evaluation of blockchain technology to create local agro-prosumer community network
- A discussion of current research gaps and design requirements to develop APCG using blockchain.

2.1 Introduction

Greenhouse gas emissions are a worldwide concern, producing an alarming climate crisis according to environmental activists and scientists. Industries, transportation, and electricity are the major contributors of GHG emissions; this is followed by agriculture and other sectors of the economy (https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions); yet agriculture and land use management act as a sink for greenhouse gases. Therefore, an imperative measure for industries is to reduce and mitigate carbon emissions. Various industries are working in this direction either by investing in carbon sequestration systems or making use of agriculture to sink the amount of carbon. However, carbon sequestration systems require substantial capital investments or advanced technology [1]. Thus, finding an effective low capital carbon capture process can be viewed as a better way to address this problem.

Since the agriculture sector can act as a sink, it is expected to use more ecological farming ways to reduce carbon emission and capture carbon simultaneously [62]. However, as the world population is expected to double by 2050 (United Nations), the agricultural sector is placed under more pressure to boost production in order to meet the increased food demand. To increase production, conventional farming makes use of agrochemicals, natural resources and fossil fuels, and operates through global supply chains which contribute to GHG emissions and creates food quality issues and security challenges. Furthermore, consumers’ food habits and food security concerns have been increasing over time. For example, consumers now demand locally grown, fresh and healthy fruits and vegetables along with sustainable environmental practices. This transitional shift brings alternative food networks (AFNs) into the forefront where environmentally sustainable practices and food quality can be performed simultaneously.

AFNs may help to meet a small percentage of the world's food demand, improve supply chain challenges and perform environmental sustainability practices simultaneously. An AFN such as the Community Agriculture Group (CSA) is a producer-consumer network started in Japan, adopted in Germany and Spain, and now a popular network in the US [23]. In simple form, CSA provides fresh local produce to consumers or members, working capital, and a market for farmers. Hence, it plays an important role in building an economic, social and sustainable network [37]. However, CSAs are highly dependent on consumers for their success, farmers' income and stability. Also, currently, community agriculture lacks a definition and structure.

With the growing urban population and limited land, other AFN forms such as urban farming, community gardens, and backyard farming are taking shape. However, the current urban farming structure shows little motivation for carbon capture. Also, it lacks prosumers' involvement and a well-defined prosumer community model which can benefit in two ways: selling vegetation and fruits, and trading in carbon credits with government bodies or industries. This chapter provides an in-depth examination of current community agriculture practices and reviews their approaches in terms of prosumer involvement, environmentally-sustainable practices, carbon farming and food traceability solutions.

2.2 Existing approaches

In the literature, very few studies have been done on community-based agriculture group involving carbon farming approaches for prosumers. In this section, we give an outline and summarize the existing research works, their approaches, limitations and gaps. An extensive literature review has been done in four major areas as shown in Fig 2.1.

Literature Review

CSA

1. Diversity and struggle[4]
2. Beyond the market[10]
3. Community supported agriculture[15]
4. Opelogicizing [19]
5. Factors impacting [23]
6. Adults and Children [30]
7. Factors influencing [34]
8. Alternative approaches [35]
9. Consumer production [36]
10. CSA in USA [37]
11. Stochastic efficiency [38]

Prosumer

1. The prosumer [3]
2. The urban prosumer [5]
3. Production for consumption [11]
4. Formation of Virtual community [14]
5. Prosumer engagement [20]
6. Pragmatic prosumtion [25]
7. Identifying prosumers energy[28]

Carbon farming

1. Impact of carbon farming[2]
2. What carbon farming [7]
3. The Australian carbon [8]
4. Carbon sequestration [12]
5. Carbon management [17]
6. Does urban vegetation [22]
7. Net carbon flux [27]
8. Soil carbon [31, 32]

Commercial projects

http://www.amazingcarbon.com/

https://regenfarmers.com.au/

https://agribusiness.io/?page_id=87

Blockchain Use in Agriculture

1. A review of Blockchain [1]
2. A blockchain frmaework [6]
3. Boundary condition for blockchain [9]
4. Blockchain based traceability [13]
5. Blockchain and more algo driven [16]
6. The rise of blockchain [18]
7. Blockchain technology [21]
8. Cryptocarbon [24]
9. Applying Blockcahin in Australian carbon [26]
10. Agriculture on the Blockchain [29]
11. Blockchain and IoT [33]
12. Emerging opportunities [29]

Commercial Projects

https://universalcarbon.com/

Figure 2.1 Literature review

The concept of a prosumer community group in agriculture is either missing or still in its early phase, and, therefore, the existing literature contains very few research contributions. Although in smart grids, prosumer community group is a popular concept [14], but in agriculture the concept is yet to be explored. Literature review is divided into four major themes: community supported agriculture, carbon farming, prosumers and blockchain use. Section 2.2.1 discusses civic agriculture and its various levels. Section 2.2.2 reviews carbon farming methods and practices in the agriculture domain. Section 2.2.3 explores prosumers in current society and examines the reasons for creating a prosumer community. Next section 2.2.4 studied blockchain use cases for food traceability and prosumers community framework.

2.2.1 Civic agriculture

"Civic agriculture", a term coined by Lyson [63, 64] refers to locally produced food and integrates it with opportunities to create jobs, encourage entrepreneurship and build strong community values. Civic agriculture works at different levels: the individual level through urban agriculture and kitchen gardens, the community level via community-supported agriculture and the state level by town planning to form socially and economically effective food systems. The garden is one of the most relaxing and interactive spaces for an urban population. "Community gardens" are places where the local community can grow plants, learn skills and meet other people. Australia has wide spaces for community gardens, which encourage urban residents to live a healthy lifestyle. For instance, in Port Melbourne, community gardens provide a space for wellbeing, social interaction, gardening and access to fresh foods. Community gardens have spread worldwide mainly for food production, income, creating employment and to improve local environments [41]. Community garden lands are usually owned by state governments, local councils, schools, churches or housing agents. There are two types of community gardens: shared gardens where gardeners work together and share what is produced, and allotment gardens where gardeners each share a small space and use it as they want (https://www.sustainability.vic.gov.au/You-and-your-home/Live-sustainably/Grow-your-own-food/Community-gardens). Community gardens benefit urbanites as they enable them to: produce their own food even if they do not own any land; exchange their cultural backgrounds; and grow traditional plants. They also help young people to learn about a sustainable lifestyle and to care about the environment. The various benefits of community gardens are listed below.

- Community gardens yield more produce per acre than traditional large-scale farming

- They contribute to better health and wellbeing.
- They can be used to recycle green waste such as tree trimmings, leaves etc.
- Community gardens are less expensive to maintain than parklands.

Another offshoot of civic agriculture is the “Community Supported Agriculture” group (CSA). CSA is a partnership between farmers (or producers) and consumers (or members) formed to share benefits, financial risk and products of farm activities [39]. CSA members issue a pre-payment to farmers for the purchase of seeds, fertilizers, farming equipment and for farmers’ efforts and services, thus easing farmers’ burden of making initial capital arrangements, and supporting their income. In return, farmers share their farm produce during the harvest period. Thus, CSA connects consumers and producers directly in a food distribution system, and commits to a safe and sustainable food supply. CSAs are attracting interest due to the current demand of food safety, locally grown produce, and sustainable lifestyles. Demand for food safety has risen exponentially and consumers now demand more and more information about food production: how a fruit or vegetable was produced, when and where it was grown, how long it has been on the shelves, how it was transported, and what processes were used to grow it organically and without any harmful chemicals. CSA safeguards food safety [35] with better food quality and taste than what is produced by conventional farming, thereby giving members a sense of satisfaction. Some big brands such as Deloitte, KPMG and PWC are providing food safety solutions. For example, PWC has developed a tool for quality assurance which works on a risk approach method to prevent and eliminate contamination. On the other hand, Deloitte is working on the future of food by utilizing new 3D printing technology, and using blockchain for food safety and food traceability, i.e., retrieving and storing product information from farm origination to logistics details and food miles and compliance. KPMG is also using blockchain and IoT to provide end-to-end supply chain traceability solutions and to offer a fast and efficient platform for better services.

Further, CSAs are economically sustained groups and do not rely on state or federal governments for capital. In fact, a CSA is not just about farming practice; it plays an important role in developing social facets such as solidarity, prosuming and de-commodification relationships among CSA members and producers [10], and guides dietary intake to encourage health improvements [65]. CSA groups develop trust between consumers and producers which helps to build sense of community. De-commodification comes through the nature of CSA groups that eliminate the total dependency of group members on the food industry or conventional farming. In addition, the mutual responsibilities of sharing capital, produce and risk, make it self-sustainable and also gives rise to prosumption.

The agricultural sector contributes to carbon emissions through food processing industries, modern agricultural practices and long supply chains. Therefore, consumers are taking steps to ensure a clean environment by shifting their focus from packaged food sourced globally to locally grown produce, thereby reducing their carbon footprints. Additionally, consumers prefer to consume vegetables and fruits grown using ecologically-sound farming methods rather than by conventional farming. In this way, they are eating healthier food and simultaneously supporting environmental sustainability. CSAs provide locally grown crops which are seasonal, fresh and better tasting than those available in conventional retail outlets. Thus, eating a CSA's local produce is a social, healthy and economic act. It also enhances the sense of community and creates employment opportunities for locals.

Farmers are attracted to using a CSA to market their produce [23]. Vegetables and fruits are the usual products shared through CSAs. Although CSAs' product offerings, the amount, season length and structure vary for different groups, their success is highly dependent on CSA members' involvement [39]. In many instances, CSA members are dissatisfied with the limited variety or the small amount of produce, which can affect their long-term relationship with farmers. For example, CSA members may at times receive vegetables that they dislike but need to consume, and often the lack of variety limits their choices. Also, in good years, there may be an excess of crops that they need to accept. Moreover, CSAs are faced with environmental risks such as extreme climate, bushfires and flood. Thus, members' continued involvement depends strongly on their satisfaction and the product quality. In addition, there are certain management challenges for CSA groups which need to be addressed to ensure smooth farming practices. For example, CSAs need to manage large amounts of crops at different stages, forecast a sufficient amount of fresh ripe produce, commit to product quality and quantity by, for example, ensuring organic farm practices and rich production to meet the pre-purchased requirements.

A major limitation of CSAs is that they do not have a similar structure and definition. In particular, a few CSA groups act as partners sharing produce and farm activities, while other CSA group members contribute only the capital and share the financial risk and produce. Currently, CSAs depend heavily on their members (or consumers). Hence, the success of a CSA relies on consumers' (or CSA members) participation and their satisfaction. Existing literature confirms this dependency and also points out the lack of a uniform structure for CSA groups.

Furthermore, these days, consumers are becoming increasingly concerned about the nature of production and its impact on the environment. For example, under its environmental and climate change program, Australia has put a price on carbon emission and created a unit to sell “Australia Carbon Unit” for compliance purposes [8]. Agriculture can implement farming practices that increase the amount of carbon in the soil. However, current CSA groups focus on agriculture and the quality and quantity of crops – they have not considered ecological farming practices that can produce carbon credits. Moreover, with the increasing urban population, community agriculture is demanding a framework for urbanites, urban prosumers or backyard farmers. However, CSAs work well for farmers with small parcels of land in non-urban or rural areas. Current CSA groups are not suitable for urban prosumers with inadequate farming knowledge or who own backyards or small garden spaces. Since the emergence of concerns about safe environment practices, locally grown crops and food security have become highly desirable. Hence, a model is needed for a CSA that is appropriate for urbanites or backyard farmers.

2.2.2 Carbon farming

Mining and construction industries, transportation, and electricity generation are some of the major industries that contribute to GHG emissions (https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions). In Australia, under the environmental and climate change program, Australian compliance carbon market has mandated that companies and other entities reduce, offset and pay for their direct carbon emissions [8]. Therefore, industries must take measures to reduce and mitigate carbon emission. Accordingly, various industries are either investing in carbon sequestration systems or making use of agriculture to sink the amount of carbon. The “Australian carbon unit” is a unit issued for compliance purposes that can be traded in with the amount of carbon offset by organizations against their carbon emission. Although the agriculture sector generates 14.5% of GHG gases, it consumes 70% of fresh water and occupies nearly 40% of the global landmass (https://www.ey.com/en_au/disruption/why-your-next-big-bet-should-be-in-food-innovation), it produces the least amount of GHG gases compared with other sectors. Moreover, agriculture and land use management are seen as potential areas to offset carbon and other greenhouse gases by utilizing carbon credit units.

Several studies in the literature discuss ways to reduce carbon emissions or sequestrate carbon through the agriculture sector. The term ‘carbon farming’ term refers to land use and land management practices that will reduce or sequestrate carbon sequestration by applying agricultural techniques that will capture carbon using a natural sink, i.e., soil and plants [7]. Previous studies have considered carbon farming methods for carbon sequestration along with various other benefits that can be both monetary and non-monetary. In regard to monetary benefits, producers or farmers can obtain economic value, i.e., employability and source of income, by selling their crops. The non-monetary benefit is that an increase in the carbon absorption rate in soil helps to sequestrate natural carbon. The amount of carbon absorbed during the crop growing process can also be traded to earn carbon credits. Thus, carbon farming not only improves the soil quality as no harmful agrochemicals are used for growing vegetables/fruits, but can also be a source of income. Ultimately, carbon farming improves environmental sustainability as it reduces or prevents carbon emissions and helps to capture atmospheric carbon. At present, the government’s changing policy and lack of awareness about the benefits of carbon credits is a drawback for carbon farming. However, carbon farming and its co-benefits are suggested [62] as a way to encourage farmers’ participation in carbon farming to boost their financial, economic and new market opportunities.

On the other hand, a comparative study has been conducted to understand the impact of carbon farming on carbon soil stock [2]. The study found a similar amount of carbon stock for organic and conventional farming techniques, and high levels of carbon formation in soil via organic farming methods. Crop rotation also helps to improve soil quality. Thus, it can be concluded that organic farming along, with periodical grazing and crop rotation, results in high levels of soil carbon.

Agricultural activities can be performed in ecological ways by avoiding tilling and modern machinery methods that emit CO_2, and refraining from using harmful chemicals in order to improve carbon absorption in the soil. However, to meet increasing food demand on a large scale, it is hard for conventional farmers to practise ecological methods. Alternative food networks play an important role here as they can feed a part of the population while maintaining a sustainable system; concurrently with crop growing, it helps to manage environmental sustainability issues such as changing climate, depletion of natural resources etc. [66]. Community agriculture, urban farming, backyard farming, and prosumption are different ways of offsetting carbon while performing ecological farming. For instance, when growing vegetables and fruits using environmentally-safe farming methods, more carbon can be absorbed in the soil, and farmers can use this to generate carbon credits or tokens and participate in the carbon market. Thus, farmers or urban prosumers (or gardeners) can earn extra revenue.

McHenry [67] proposed a pyrolysis process to utilize bio-char for production and the carbon sequestration process. Bio-char (biologically derived charcoal) is suggested as another option that can improve agricultural production, increase natural rates of carbon sequestration in soils, reduce farm waste, and substitute renewable energy sources for fossil-derived fuels. However, carbon sequestration methods alone result in less carbon content in soil when compared to changes in land use management and improved overall agricultural practices. Carbon sequestration processes alone often increase fossil fuel use such as electricity, which adds to carbon emissions. A number of studies as shown in Fig. 2.1 have been carried out on the use of advanced technology and methods for carbon sequestration, which indirectly adds to carbon emission or disturbs the natural phenomenon of carbon capture. So far, carbon farming is a natural and sustainable approach that can benefit farmers as well as proving monetary and non-monetary value.

Hence, carbon farming practices along with community agriculture can be combined to take advantage of the benefits of both. Since carbon farming requires motivation, support and understanding of government policy and awareness of its co-benefits, the establishment of a community of farmers or urban prosumers can facilitate carbon farming by promoting the group, providing motivation and incentives, and a platform for sharing/selling produce and carbon. Hence, in order to get benefit from carbon farming, prosumers or urban gardeners can form a community to gain better value from carbon tokens and market exposure while working together.

2.2.3 Prosumers

Prosumption is an old work culture dating back to when people used to grow their own vegetables and fruits. Toffler coined the term "prosumer" to describe production that is undertaken for self-use. Although industrialization era introduced market where producers and consumers separated from each other, prosumption is making its way back through the demand of sustainability and safe environment (https://www.scp-centre.org/promona/). "A prosumer is a producer and consumer in one person" [5]. In modern times, prosumption is not a necessity; rather it is a choice made by people who are concerned about food safety and a sustainable environment. The main reasons for becoming a prosumer are related to health, and saving money and time. Because prosumers' crops are generally home grown and organic, they are considered to be healthier and better tasting than those available in supermarkets. Also, prosumers seek to save money by growing their own vegetables and fruits. Other factors which are attracting interest in prosuming are discussed below.

a. Demand for locally grown produce: Agriculture supply chains are complex and risky, as they involve various intermediaries, volatile factors such as weather, disease, storage etc. which makes produce costly, insecure, poor taste, or quality that may cause ill-health. Nowadays, people are more concerned about what they eating, how and where it comes from, whether crops have been grown organically, or whether they are locally produced. Hence, there has been an increased demand for local food growers and suppliers. Consumers are shifting their focus from global organic produces to locally grown organic produces due to ecological, environmental and cultural concerns. Locally produced food shortens the long, complex supply chain, supports the economy and, generally, the produce is fresh and better quality. Hence, prosumption is the solution for locally grown vegetation and fruits, which is limited to a certain geographical region and also adds to social and cultural sense of community. Consumers are also engaged in production by becoming members of a community-supported or DIY agriculture group.
b. Food safety and quality: Consumers are concerned about security of supply, supply chain efficiency, and trust between the actors and entities. Trust and safety are inherent in prosumers' crops as they are home grown, and sharing and distribution do not rely on a long and complex supply chain. With increasing food-borne illnesses and food recalls, consumers expect to know the quality and safety of their food. Prosumers' participation in production allows comparatively easy food traceability and better-quality produce. Generally, the quality of prosumers' crops is more trustworthy than that of commercial producers.
c. Environmental sustainability: With the excessive amount of carbon dioxide in the atmosphere, environmentally-friendly substitutes of our activities are coming into play. Prosumers help to create environmentally-friendly solutions by cultivating vegetation in ecological ways by, for example, using compost, avoid tilling and harmful fertilizers, and producing their own fossil fuel. 'Carbon gardeners' is a term used to encourage carbon sequestration practices in one's own backyard (garden) or by forming a group (https://www.ecologicalgardening.net/2015/09/why-not-start-today-backyard-

carbon.html). Prosumers' crops are usually organic, limited to small scale production and grown in healthy soil. Thus, a high amount of carbon is stored in the soil, which gives high quality produce and helps with carbon sequestration.

d. Economic and social integrity: The faults in the current economic system suggest the need to create self-provisioning and alternative economies. Self-provisioning individuals are those who eat their own home-grown food and do not rely on the food industry [11]. There has been an increasing recognition and appreciation of prosumers and their goods and services. Prosumers enjoy a sense of community where they form social relations, exchange knowledge and learn skills from each other. While producing for self-use, prosumers create a self-sustained environment.

Moreover prosuming either individually or as a group is seen as a political act which is feasible, reduces unfavourable environmental change and affects the economy by reducing centralised, long supply chains [5]. Prosuming activities are also seen as meaningful, recreational and providing good social relationships. Performing tasks with your own hands not only increases skills, but also improves knowledge which helps to develop creativity and improvisational skills. Additionally, on the personal level, prosuming promotes self-efficacy, empowerment, self-sufficiency and pride about producing your own food [5]. On a practical level, prosuming supports the household and its resilience. Protecting the environment is seen as an overall value. Thus, prosuming in urban agriculture can be perceived as a sustainable step for the agriculture industry. However, currently, there are various limitations to the prosumer model. For instance, prosumer activities are time consuming, require motivation and often expensive. The prosumers have to deal with legal hurdles because of the absence of a formal structure. Moreover, new prosumers who want to join often lack knowledge and skills. Hence, there is inadequate prosumer management, insufficient information about and definition of prosumers; hence the gap in the research.

A prosumer has good prospects of obtaining better value and exposure in the market by joining a prosumer community group. The formation of a community prosumer group can also address the shortcomings of prosumers. To date, not many studies have been carried out to define a prosumer community group and propose a detailed structure.

To overcome prosumer food safety and trust issues, a central management body can be established. However, central systems again have trust and traceability issues and require intermediaries. Hence, identifying and using an advanced technology which facilitates food traceability and is decentralised seems a solution. Blockchain is a popular technology used to build trust-less systems, and therefore can provide a solution.

2.2.4 Agriculture using blockchain

With the current desire for transparency and traceability in the agriculture and food industries, it is important to use a technology which is unmodifiable, transparent and able to store information securely. Blockchain, a distributed ledger technology, is an innovative approach to building decentralized trust-less systems. Various businesses are building solutions for agricultural and supply chain issues via making use of blockchain technology or concepts [68]. IoT and blockchain are combined to develop solutions for agriculture traceability problems. Brewster et al. [69] discuss the important role of IoT technology and culture adaptation to deal with agriculture using IoT-based, large-scale pilots. Caro [13] presented the AgriBlockIoT system to integrate blockchain and IoT technologies to solve the agriculture traceability problem. A large number of studies have been done to utilize blockchain for food safety and traceability issues of agriculture [21, 70]. Blockchain use cases are also examined, which include agriculture as one of the major sectors [71]. Currently, various projects have been initiated to solve agriculture issues such as supply chain, traceability, food security, food safety and carbon market using blockchain technology [13, 21, 24, 26, 70]. There are certain limitations and challenges associated with blockchain [18] such as the complexity and immaturity of blockchain technology in terms of its practical application in the agriculture industry [1] . Also there are issues related to cybersecurity in technology based logistics solutions [72].

Other than transparency and traceability issues, global supply chain and large-scale agriculture systems affect small-scale producers or alternative food networks in the market. In particular, small producers or prosumers experience difficulty in entering the local or national markets. Prosumers and small-scale farmers struggle to access finance, and face issues of access to markets and transportation. Also, small producers find it difficult to match the range of quality standards, traceability and certifications [29]. A smooth and trusting connection between prosumers and the consumers or market can generate better growth and income for prosumers and small-scale producers via AFNs, which in turn can contribute to meeting the increasing food demand. To date, blockchain has been used to solve global supply chain issues, food security and food traceability complications. However, there is little experimental research on the application of blockchain technology for the purpose of creating a prosumer community structure for better market, connection and income for small-scale producers, urban farmers and prosumers.

2.3 Shortcomings

The comprehensive study of the literature comprised four major areas which were examined in detail: community agriculture, carbon farming, prosumers and blockchain. The literature review revealed that the increasing demand for food and the changing climatic conditions are calling for sustainable alternative food network such as community agriculture or prosumer community groups. So far literature review shows no prosumer community sharing network framework is discussed in agriculture scope. Also current community agriculture has limitations such as lack of structure of community group, criteria for new members wishing to join, motivation to work in order to create a sustainable environment and carbon assessment framework for farmers. And little consideration has been given to urban population/backyard farmers in the context of community agriculture. These limitations are discussed in detail below.

2.3.1 Lack of uniform detailed structure for community agriculture groups

Community supported agriculture are popular consumer-producer groups. Although current community supported agriculture relies heavily on CSA members' participation and high levels of satisfaction for its success [73]. Additionally, current CSAs do not have a uniform model or detailed structure which often compromises farmers or producers to work for a modest income to very modest income, or engage new farmers [73]. There also seems inequality of land and labour and shortcomings in the current CSA structure, which demands a new theoretical proof of concept of a framework that can solve production cum environment issues [74].

2.3.2 Lack of urbanites and prosumers participation in alternative food network

With a steadily increasing world population, the demand for alternative food networks (AFNs) such as community supported agriculture groups, urban farming and prosumption, are increasing. Although AFNs contribute to a small percentage of food production, it is expected that they will continue to grow and meet the increasing demand for local food. Currently, community-supported groups work well for large-scale agriculture systems or small-scale farmers. However, there is little participation in CSAs by the urban prosumers and consumers [39].

2.3.2 Lack of definition for new prosumers to join

Backyard farmers or prosumers are emerging as another prospect for community agriculture. However, current community agriculture groups see farmers as producers, and consumers as CSA members and utilizes acres of land [75]. A small producer (who produces in a small space such as backyard, garden or balcony) or individual with less agricultural or farming knowledge, seems unsuitable as a farmer or producer in CSAs.

2.3.3 Lack of sustainable structure for current community agriculture practices

The main concern of the agricultural sector is its ability to meet the demand for produce. However, to increase the yield, conventional farmers often apply harmful chemicals and fertilizers to boost production, which in turn increases the carbon emission. In addition, advanced machines and techniques such as tilling, use fossil fuel, while garden waste and electricity produce more carbon emission. Although agriculture and land management are considered as a natural sink to offset carbon, ecological community agriculture practices can contribute to sustainable environment. The government has created carbon credits to offset carbon emissions, which is a practical way of earning additional income, aside from earnings obtained by the sale of produce. Since farmers are often not motivated to practise ecological farming or are not aware of the policy or benefits of using it [8], a gap exists here created by a lack of common objectives or the motivation to engage in carbon farming as a community agriculture activity.

2.3.4 Lack of prosumer community agriculture group

Community agriculture consider a partnership of farmers and consumers. However, prosumers who produces and consumes both often find prosuming activities time consuming and expensive [38]. Moreover, they have to deal with the legal issues and may not be able to cope with administrative tasks [34, 76]. Prosumer community groups are ideal for prosumers who want to get better value and market. Additionally, recent studies have paid little if any attention to prosumer community groups and their structure in the context of agriculture.

2.3.5 Lack of transparency and trust in community agriculture group

Blockchain is widely used in the agricultural sector to solve transparency and trust issues. However, blockchain is not being utilised for the creation of community agriculture structures for better markets, connections and income for urban farmers and prosumers. Also, in regard to carbon trading, various companies are inclined to use blockchain, although it has not been applied to community sharing networks, or as a platform for carbon and produce transactions.

Thus, a formal structure or a new model is required for the establishment and control of community agriculture group activities. Also, it is essential to develop a model that provides a platform which can help build a self-sustained, self-regulated and transparent prosumer community group to share produces and carbon. As a solution, a framework is required for the prosumer community group that provides a seamless structure for prosumers definitions, connects new prosumers to the group, motivates them, and estimates the carbon absorption amount to trade with outside entities. The literature review showed the need for a proper structure that is better managed and can become self-sustainable and self-regulated. The model must offer transparency and, most of all, ensure food production in a safe and decentralised manner.

2.4 Design requirements of this framework

In the previous sections, a comprehensive literature review related to existing community agricultural group approaches and their shortcomings is performed. A number of technical challenges have been identified post review, which need research attention. Therefore, in this chapter, the focus is on the challenges or limitations found in current community agriculture systems concerned with a sustainable environment. Below are the major limitations.

- How to define APCG concept?
- How to engage new prosumers to the existing prosumer community groups?
- How to motivate the prosumers to commit to the sustainable environment cause?
- How to maintain trust in prosumer networks, i.e., between prosumers and the market?
- How to maintain an economically self-sustained group?
- How to ensure food traceability/carbon token transaction?

2.4.1 How to define agriculture prosumer community groups?

The creation of meaningful APCGs is the first step in establishing a sustainable agro-prosumer community group (APCG). In order to do this, a framework is required that defines agro-prosumers and identifies their production characteristics. These production characteristics will then be used as pre-requisites for each group and will be applied to new prosumers wishing to join the relevant APCG.

2.4.2 How to engage new prosumers into the existing prosumer community groups?

The ongoing engagement of new prosumers is important in a growing APCG network. In order to achieve sustainable prosumer recruitment engagement, it is essential to analyse prosumers' production and consumption behaviour, types of vegetation and fruits, age of tree/plants, and the type of produce (i.e., organic, inorganic). Recruitment seems complicated when prosumer does not have a historic profile and experience in agriculture. Analysis will help in identifying eligible and non-eligible prosumers.

2.4.3 Motivating prosumers to engage in sustainable environment practices

Prosumers can undertake carbon farming by becoming carbon gardeners. By applying ecologically-sound methods of farming, the total amount of carbon content in the soil can be increased via plants/trees. Prosumers can sell the total amount of carbon offset, obtained by growing vegetables and fruits, to industries or government entities and earn an Australian Carbon Credit Unit (ACCU). In this way, participation in a prosumer community group can lead to both monetary benefits (i.e., income from the sale of crops and carbon credits) and environmental benefits (i.e., healthy soil, healthy and organic produce and carbon reduction). Thus, a framework is required to assess carbon capture, which can be traded with external organization to offset their carbon emissions, thus improving the environment.

2.4.4 How to maintain trust among prosumers and market?

Trust is an important aspect of any relationship. However, it is challenging these days to build and maintain trust in a network which has intermediaries and is operated by a central system.

In agriculture consumers are more interested in knowing the origination of produces, who are the producers, method used by them etc. which is often lost or manipulated in long supply chain or in global market [16]. Hence, an advanced technology is required to build a trust-less model.

"Blockchain is a distributed ledger, shared across a public or private computing network". The main advantages of blockchain are that it is decentralized, immutable, transparent and secure [77]. Hence, it allows information to be verified without involving a third party. Therefore, APCG is proposed using blockchain concepts, where food traceability and recalling can be achieved.

2.4.5 How to maintain economically self-sustained group?

Blockchain technology provides a "smart contract" to record rules and regulations and verify every transaction against the smart contract conditions and is triggered only if the conditions are met. Thus, blockchain can be utilized to simulate the framework prototype.

2.4.6 How to ensure food traceability/carbon token transaction?

Food traceability is an important factor for the safety and quality of agricultural products. Transparency is also a desirable feature for carbon trade. Hence, blockchain is used to accommodate these features. APCG solves food traceability issues by shortening the long supply chain, and with the help of blockchain, a smart contract feature can be used to trace prosumer id on each product for easy tracking.

This thesis focuses on four key challenges associated with the creation, purpose and management of APCGs for agro-prosumers. The challenges are: define APCG concept, how to engage new agro-prosumers, how to manage diverse goals in a community-based approach, and how to measure carbon capture and implement APCG on blockchain. Finally, the model will be implemented using blockchain to maintain trust and sustainability.

2.5 Conclusion

This chapter has offered a state-of-the-art review of the management of agro-prosumers through APCGs, as well as the literature pertaining to prosumer community groups, carbon farming and blockchain in the context of produce-sharing networks for a sustainable environment and to address local food demand and traceability issues. The literature indicates that the notion of agro-prosumer community groups is still a fairly new concept. Although the concept of a prosumer community group is well established in smart grid, it is still new for the agriculture sector. For urban farmers or prosumers, a community group established for sharing produces/carbon seems highly attractive because of the growing demand for locally-produced food, food safety concerns, and environmental issues. Prosumers' participation in a community group has various benefits which can be monetary (income from the sale of produce) and non-monetary (environmental benefits derived from carbon capture by carbon gardeners/prosumers).

Chapter 3

3 Agro-Prosumer Community Group (APCG) concepts

This chapter provides

- An introduction to the APCG concepts
- A framework to define agro-prosumers and categorise them into APCG groups
- A conceptual process of outlining proposed framework requirements, design logic and theoretical base.
- Verification of framework using prosumers dataset.

3.1 Introduction

Prosuming, individually or as a group, is seen as a political act which is feasible, reduces unfavourable environmental changes and affects the economy by reducing centralised long-chain value in the supply chain [5]. Thus, prosuming in urban agriculture can be perceived as a sustainable step for the agriculture industry.

With the growing interest in the supply of quality food, the urban prosumer can be seen as a leader in providing high quality, trustworthy produce. Prosumers' crop yields are high quality as generally they are home grown, chemical free, grown in better soil (i.e., composting) or grown organically. Additionally, prosumer crops are more trustworthy and better quality than commercial produce as they are usually grown from seeds or seedlings, as is the practice of the urban prosumer.

An agro-prosumer has strong prospects of obtaining better value and exposure in the market by forming or joining a prosumer community group. A agro-prosumer community group can improve economic value and offer rich socio-psychological experiences [5] to the agro-prosumers by creating social relationships and self-pride, imparting new skills to members, generating knowledge, and contributing to community activities. In addition to high quality produce, agro-prosumers can generate carbon tokens depending upon the consumption of the total amount of carbon content consumed during the vegetation process, and trade it with industries.

An agro-prosumer is defined as a producer or grower, who produces vegetation/fruits in its garden (backyard) and vends the surplus to the interested buyers. An agro-prosumer community group (APCG) can be described as a community group formed by “using different agro-prosumers profile, personal motivations and unique characteristics”. As discussed in previous chapters, a framework that defines agro-prosumers, identifies their characteristics and interests to categorize them into groups and finally engages them via a goal oriented network will not only create an alternate food network but also solves carbon sequestration problem to a certain extent. Hence this chapter proposes a smooth structure of APCG framework, which contains four sub frameworks i.e. define agro-prosumers, recruit new prosumers, goal management and carbon assessment.

The first component of APCG is to categorize agro-prosumers into groups based on their characteristics and production behaviour for better performance and operations. The main objective of this chapter is therefore to define the agro-prosumer community group and determine its features to derive real prerequisites which will be utilized to categorize and cluster different agro-prosumers into groups.

The data presented in this chapter is feed into database which will be used to construct a carbon token/production-sharing framework designed using blockchain technology in later chapters. This chapter is organized in following way; an outline of APCG concept is given in section 3.2. The requirements of the framework and the design logic are presented in sections 3.3 and 3.4, respectively. Theoretical fundamentals of the proposed framework and its algorithms are covered in section 3.5. Section 3.6 shows the verification of the framework, followed by the conclusion in section 3.7.

3.2 Outline

APCG takes prosumer’s produce profiles as a key input. Agro-prosumers’ produce profiles define agro-prosumer which includes their information, postcode, property_id and types of crops grown in their garden, along with the quantity during different seasons, particularly winters and summers. The framework is shown in Figure 3.1.

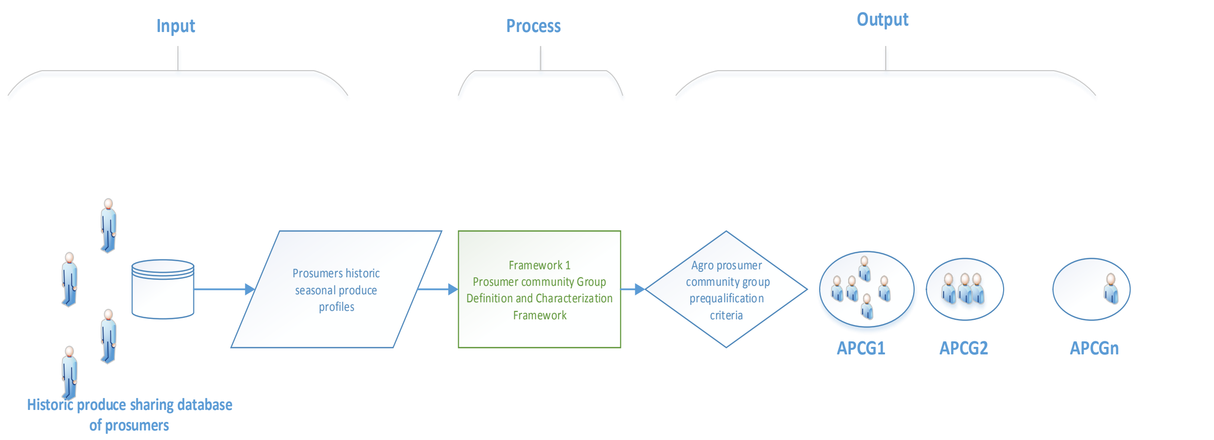

Figure 3.1 An agro-prosumer community group definition and categorization framework

The framework comprises a three-step process: (i) the agro-prosumer's profile from the database will be reviewed and selected by the selection module; (ii) the selected profiles from step 1 will be grouped according to postcode, and clustered in geographical based cluster groups according to their proximity; and (iii) for each geographical group, the corresponding agro-prosumers' "quantity and types of vegetation/fruits grown" are analysed to create agro-prosumer community groups (APCGs). Finally, for each geographical group, clusters are formed which will be represented as APCG and analysed for unique charactestics.

In the first step, the selection of the profile is based on the postcode and highest yield count (HYC) and lowest yield count (LYC) for fruits or other crops as shown in figure 3.2. Agro-prosumer profiles will be processed to identify suitable profiles for the APCG, and reject the rest. In the second step, postcode is determined and the profiles of prosumers in the neighbourhood are grouped together for easy delivery. Lastly, a combination of HYC and LYC, and vegetation type will be used to identify each group's pre-requirements for membership.

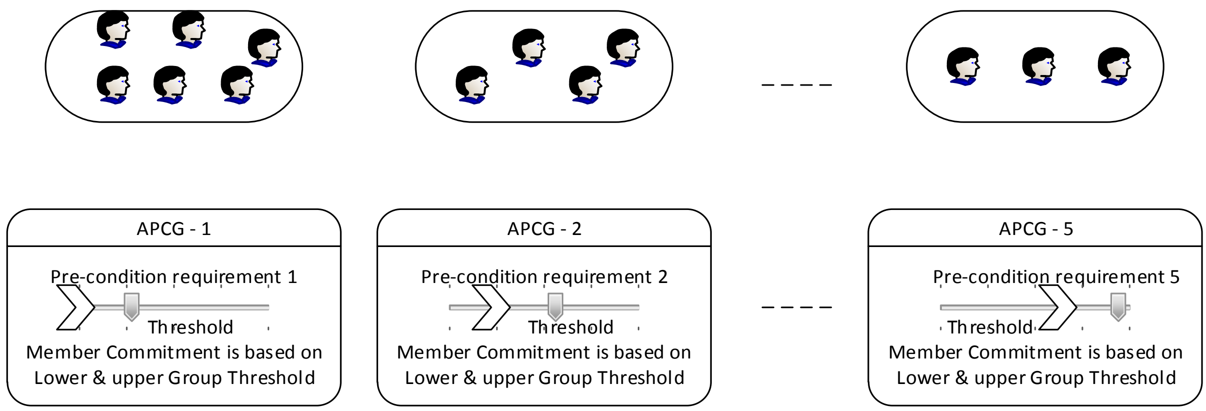

Figure 3.2 Pre-requisites of APCG

The pre-requirement for each member will be treated as a commitment to meet their group's prerequisites. For this research, lemon trees and plants are focused to build the framework.

3.3 Requirements

For the proposed framework, the first stage is to obtain agro-prosumers' produce profile. Following are the requirements for the produce profile selection process:

- Collecting agro-prosumers' produce profiles: To define an APCG, registered produce profile details such as address, postcode, seasonal produces, production quantity, age of a tree, height and diameter of tree, number of trees, land size, type of soil, organic or inorganic ways of farming and number of persons in a family are collected. The values can be extracted through observation either based on prosumers' past history or studied over an evaluation period.
- Identification of suitable produce profiles for APCGs: Based on the attributes extracted from the produce profile, profiles with extreme values or distant postcodes are removed. For example, the observations which are significantly removed from the mean of the data values will be considered as outliers. Finding outliers and excluding them will help in identifying suitable profiles for the framework.
- Creating agro-prosumer group on the basis of similar pattern: One major part of the framework is to make sure that similar pattern profiles are chosen when forming a group to facilitate management and ensure similar entitlements and benefits. In the agriculture industry, clustering can be done on the basis of postcode to facilitate logistics activities such as storage and delivery. Later location specific groups can be created using similar production behaviour.
- Attaining the feasible number of APCGs: The aim of the framework is to form groups that have maximum efficiency and effectiveness. This can be achieved by ensuring a reasonable number of agro-prosumers in a group. For instance, a large number of agro-prosumers in a group may over produce, which in turn can present storage issues or damage the production. Conversely, a small number of agro-prosumers in a group might not be able to meet the demand, particularly if one or two members do not produce the expected quantity.
- Define prerequisite for each group: In order to have a sustainable framework, each prosumer group should clearly classify produce profiles into different APCGs based on

their production behaviour, thus enlist APCGs entry requirements. These pre-requirements can be referred at the time when a new prosumer wants to join APCGs.

3.4 Design Logic

A set of design logics are presented here to fulfil the requirements discussed in the previous section. The design of the proposed framework is shown in Figure 3.3.

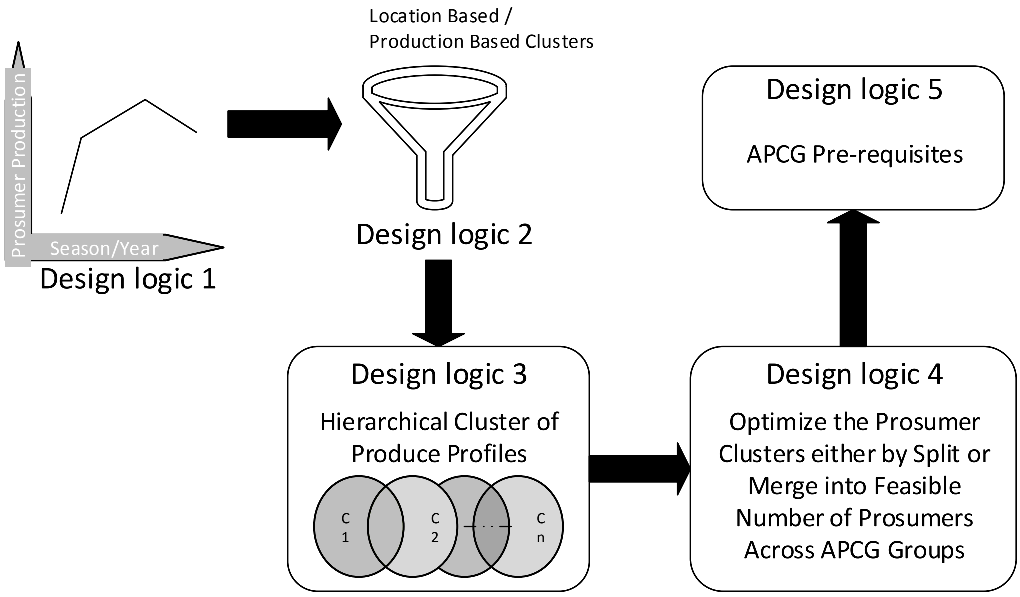

Figure 3.3 Outline of proposed framework

Design logic 1

The first step is to determine the postcode of individual prosumers and their produce. Then, other information is collected: number of family members, lemon-growing data such as records of lowest and highest yield count based on past produce history or age of plant, and type of farming and amount of consumption. Agro-prosumers' production behaviour are obtained using their seasonal produce profile. Every agro-prosumer is defined for two climatic conditions (i) seasonal produce profile for summer; (ii) and seasonal produce performance for winter. Annual production is considered if the crop is produced all year. This framework pertains to lemons for the purposes of this study; therefore, different lemon varieties such as Eureka, Meyer and Lisbon are considered here.

Table 3.1 Production profile of the prosumer

Season	Produce performance
Winter	Data from March-August (~6months)
Summer	Data from September-February (~6months)
Yearly	Data from January- December (~12months)

Design logic 2

Design logic 2 is used to identify suitable agro-prosumer produce profiles in order to define different APCG clusters using clustering techniques (machine learning unsupervised algorithm). The prosumer profiles are partitioned into groups in two phases.

- Location based clusters: In the first phase, the prosumers are grouped into clusters using their postcodes for a particular region. This rationalizes the delivery process in each cluster since prosumers in each cluster can market their produce within a minimal area.
- Outlier detection: The produce profiles that are not from majority of suburbs (or far distance) are detected as outliers; i.e., the data points are far from the rest of the data.

Design logic 3

Since hierarchical clustering is performed to create clusters based on the homogeneity, it is used to perform the categorization of prosumers. Reason to select hierarchical is simple as we want to cluster similar production behaviour prosumers in a group. Clustering based on similarity will help in keeping fair setting for all prosumers. For instance, the production expectation quantity, incentives/profit distribution and sustainability can be achieved adequately. Thus, similar produce profiles will be grouped according to the proximity of prosumers. This step satisfies the requirement for agro-prosumer clusters to be based on the similarity of agro-prosumer profiles. The output of this step will help to determine suitable pre-qualification criteria for entry into APCGs.

Design logic 4

Once clusters are formed, they be too large or too small to be successful. Hence, to maximize groups' efficiency, small groups are merged or large groups are split, again using the similarity behaviour to determine proximity. This will ensure that each group has sufficient numbers of agro-prosumers and produce to work efficiently.

Design logic 5

The prerequisites of each APCGs is defined by the attributes of APCGs discussed in the previous steps. As derived initially, the seasonal produce will be obtained and examined to derive "LYC" and "HYC". LYC will be kept as the prosumer's entry requirement for its corresponding group, and HYC will become the prosumer's maximum accountability to its corresponding APCG.

The above design decisions are used to build the proposed solution framework for APCG concept. The next stage involves the theoretical basis for the proposed solution.

3.5 Theoretical Fundamentals

Next stage of the conceptual process is the theoretical fundamentals for the APCG concept framework as shown in Figure 3.4. The fundamentals are presented in two phase - Phase 1: clustering prosumer profiles and outlier detection, and Phase 2: optimising prosumer clusters and defining group pre-requirements.

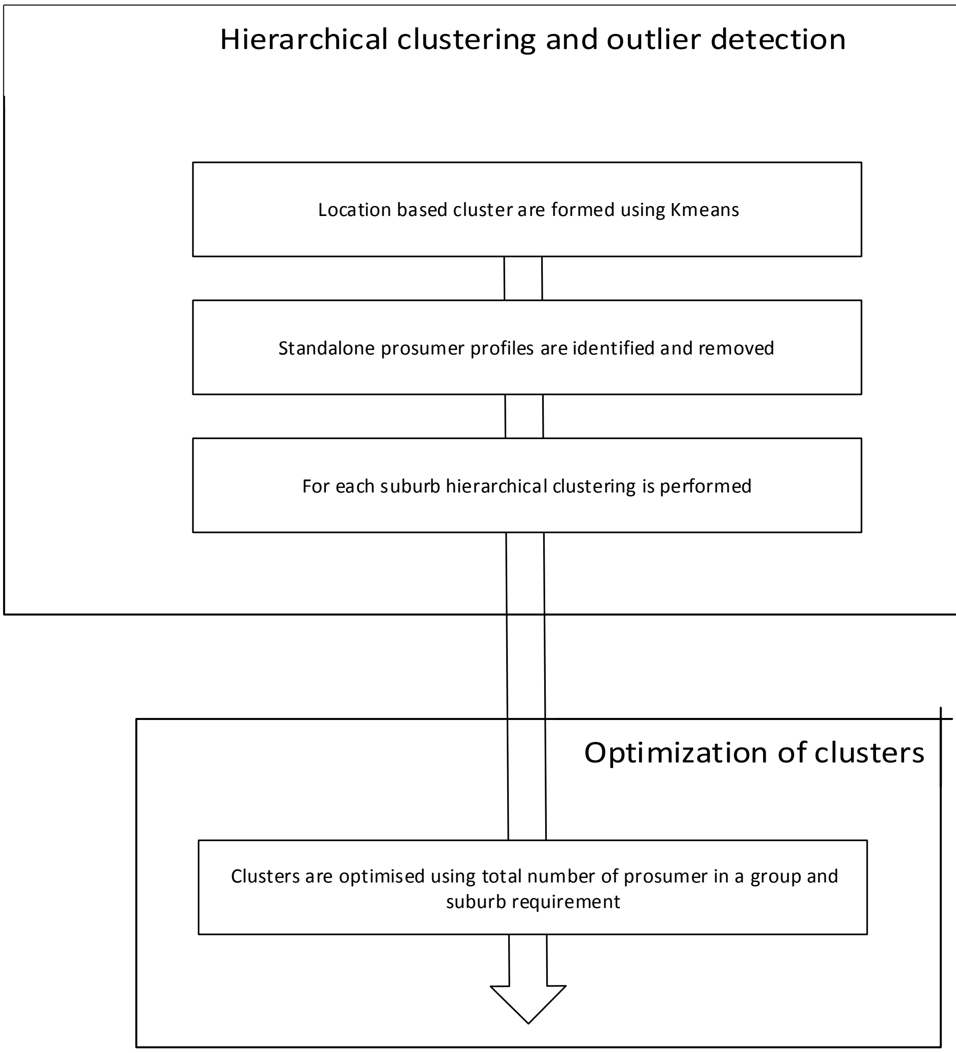

Figure 3.4 Theoretical fundamentals of APCG

Agro-prosumers' seasonal summer and winter data has been collected as an input for the first part of the framework. A hierarchical clustering algorithm shown in Fig 3.5 is used to create clusters based on the homogeneity of prosumers' profiles.

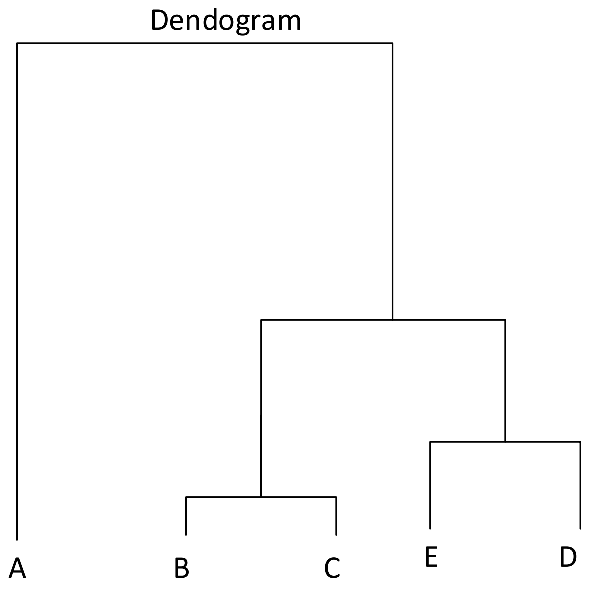

Figure 3.5 Hierarchical clustering

Clusters are optimised and unique attributes are identified in the second part of the solution. Clusters that are not overlapping are then optimised to achieve a feasible number of APCGs in next phase. Additionally unique attributes are identified for each group and act as pre-requisites for APCG groups.

The next section provides a detailed description of the proposed theoretical development.

3.5.1 Phase 1: Prosumer Cluster-grouping

The first phase includes clustering of the prosumers' profiles using a hierarchical clustering method. Prosumers' seasonal profiles for two seasons (summer and winter) are taken as an input for this phase. There are three steps in this phase: creation of regional groups, building clusters, and outlier detection.

3.5.1.1 Step I Creating regional groups

In this phase, agro-prosumer's postcode is taken as an input. The prosumers are partitioned into groups based on their postcodes within a certain region. This will mean that the delivery of prosumer produce can be done without the need for long-distance transportation. The output of this step will provide GL-clusters (geographical location based- clusters) based on postcodes and the neighbourhood zone.

3.5.1.2 Step II Outlier detection

In order to deal with outliers, a threshold is set: after calculating the distance between existing clusters, if the shortest distance is not further than the threshold, we assign the dataset to its closest cluster. If the shortest distance exceeds the threshold, this means that the cluster could belong to a minor group. The objects in the minor group are those that did not belong to any major groups. Objects in minor groups are data points, not outliers as they do not belong to any major groups. Further clustering of objects in minor groups can be done for future analysis.

3.5.1.3 Step III Building clusters

For each GL-cluster obtained from step one, the corresponding agro-prosumer profiles are considered for the next step, and clusters are formed based upon prosumers' production history. The hierarchical method is used to decide the number of clusters. In this process, initially, each prosumer profile is placed in a unique cluster. For each pair of clusters, some value of dissimilarity or distance is computed. In this case, minimum variance, i.e., "ward's criterion", is used to minimise the total within-cluster variance and find the pair of clusters that leads to minimum increase in total within-cluster variance after merging. In every step, the clusters with the minimum variance in the current clustering are merged until the whole dataset forms a single cluster. Hierarchical clustering helps in identifying groups in the dataset. Thus, the output from this step will be number of prosumer clusters based on their production similarity.

3.5.2 Phase 2 Prosumer Cluster Optimisation and Forming Pre-requisites

This phase involves the optimisation of prosumer groups based on the number of prosumers in each group and their production amount. Firstly the clusters are optimized and pre-requisites for each cluster-group is formed. The optimization steps and pre-requisites are further illustrated in this section.

3.5.2.1 Optimisation of cluster-groups

Agro-prosumer cluster-groups created by using hierarchical clustering are optimised to produce sufficient number of clusters that will then represent different agro-prosumer community groups. The number of clusters produced by optimization, depends on the variation of production quantity. If the variation is large, too many clusters could be formed, which are not feasible to manage. Thus, this stage involves optimising the clusters into a feasible number of APCGs by merging small clusters into one or splitting large clusters into smaller ones to obtain a feasible number of APCGs to satisfy market requirements. In order to determine the ideal number of clusters, firstly, suburb requirements are analysed and the expectations of relevant APCGs are derived.

To optimise APCGs;

Let X be the population of suburb ABC and C is the per capita consumption of lemons. Assume that the APCG framework targets a minimum 1% of lemon market for a suburb ABC. Then the requirement ($R_{expected}$) of lemons for suburb ABC using APCGs can be calculated with

$R_{expected}= X*C*0.01$

Let L be the number of clusters formed using the clustering method, and R_L represents every APCG's goal.

$R_L= R_{expected}/L$

After determining the suburb's requirements, next step evaluates number of agro-prosumers present in each APCG (as shown in Figure 3.6). Let say P_l and P_h respectively be the lowest and highest number of prosumers expected in each group. Let R_L be the minimum amount of production expected from each APCG. Prosumers count (P_{num}) and the production quantity ($R_{obtained}$) from a specific prosumer cluster is shown in equations 3.1 and 3.2.

$P_l<P_{num}<P_h$ Equation 3.1

$R_{obtained}>R_L$ Equation 3.2

If the production is less than the expected amount ($R_{obtained}<R_L$), or the number of prosumers is lower than the ideal number of prosumers ($P_l>P_{num}$), that agro-prosumer cluster is merged with the closest prosumer cluster, and the same process continues until the prosumer cluster can meet the total production requirement ($R_{obtained}>R_L$) and the number of prosumers ($P_l<P_{num}<P_h$) defined for the APCG.

Now, if too many prosumers form an agro-prosumer clusters, the clusters are further break down into small size clusters consisting of an most favourable number of prosumers and meeting the production goals i.e. $R_{obtained}>R_L$ and $P_l<P_{num}<P_h$.

The final output of optimization will result in the optimised prosumer clusters, which are then represented as APCGs. Now these APCGs are analysed to identify the unique production characteristics or pattern of each group which will be denoted as the pre-requisite of the APCGs.

3.5.2.2 Pre-requisite formation

Introduction to APCGs includes formation of unique entry requirements for each group. The two key input, as discussed in the previous sections, to determine the prosumers' adherence are the "lower threshold" (*Lt*) and the "upper threshold" (*Ut*). The defined inputs are:

- Lower threshold: L_t
- Upper threshold: U_t

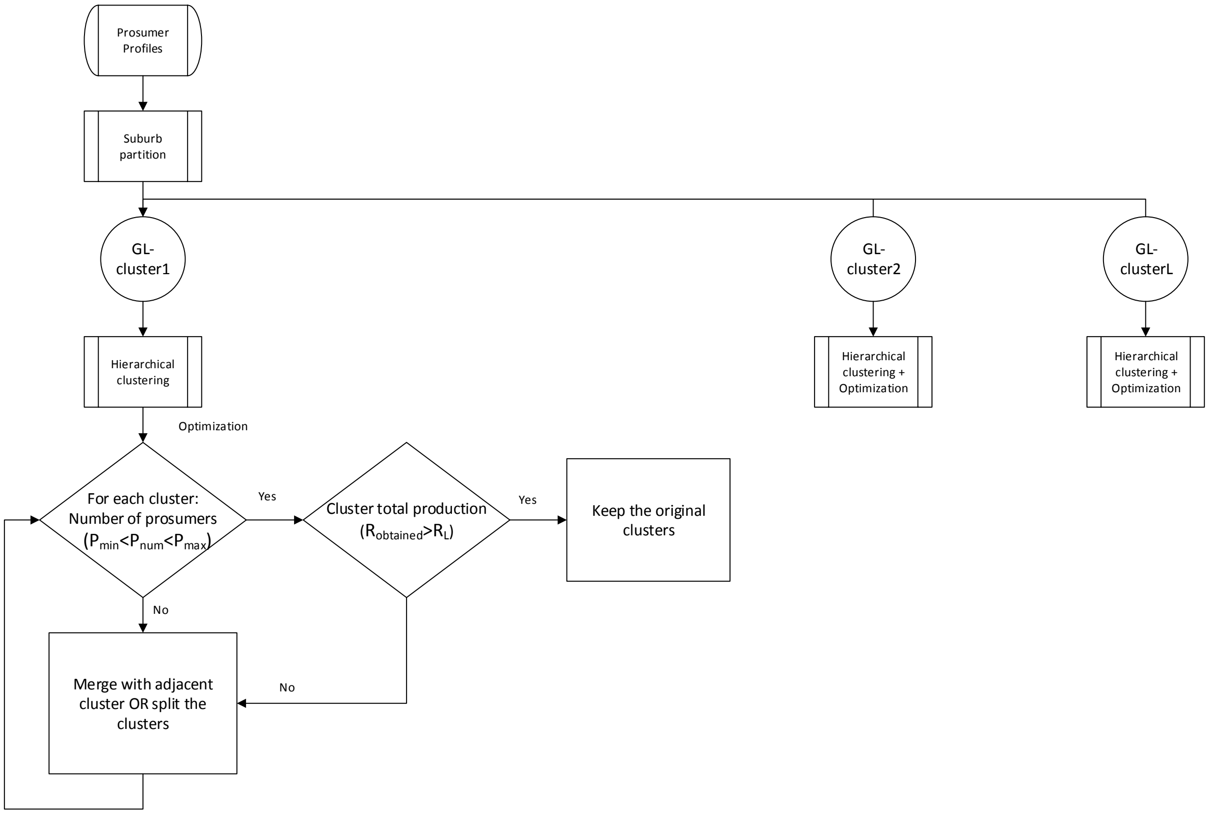

Figure 3.6 Flowchart for Hierarchical and Optimization of clusters

3.6 Verification of APCG concept

Prosumer parameters are illustrated for the verification of the framework. For verification purpose, R software and programming language have been used. R is a popular tool for statistical computing, programming and graphics. R is used to scale datasets, develop machine learning algorithms and build models and applications. In addition to R, traditional Excel is used to prepare and analyse data. Next a prototype is created to verify the proposed framework.

3.6.1 Prototype

In this section, we illustrate the prototype set up and the parameters used for the verification. The following parameters are used for prototyping the APCG concept framework.

Table 3.2 Prosumer prototype parameters to verify APCGs concept

Prototype parameters	Value
Numbers of prosumer profiles	200
Minimum and Maximum threshold distance for outlier detection	2km-10km
Minimum agro-prosumer participants in a group Maximum agro-prosumer participants in a group	10 50
Minimum accumulated lemon production expected from each APCGs Maximum accumulated lemon production expected from each APCGs	50 2000

Table 3.2 shows the key parameters for the verification are the prosumer production dataset. This framework is proposed using one type of crop only: lemons. It is challenging to obtain a dataset for lemon yields because prosumer community group data is not publicly available yet. Therefore, prosumer production profiles are generated using minimum and maximum lemon production and consumption. In the sub-section below, we discuss the generation of prosumer profile data.

3.6.1.1 Prosumers' production dataset generation

In this section, prosumer profiles are generated using the Australian standard production and consumption pattern (as shown in Table 3.3).

- Country/region: In order to generate prosumer profiles, production parameters are analysed particularly for the State of Victoria, Australia. For this study, prosumers residing in Victoria are used only to generate a sample data set. Therefore, Victorian suburban postcodes are randomly generated for prosumers. The average residential block of land is utilized to generate land sizes across Victoria. For each postcode, latitude and longitude values are determined in order to build prosumer community groups that are in close proximity.

- Vegetation/fruit: Lemon trees generally produce the first crop after three years, and reach maturity when they are about five years old. Hence, the age of lemon trees and the variety are considered when estimating the minimum and maximum number of lemons produced during harvest season, and assessing the amount of carbon absorption. For this study, we consider three of the most common varieties: Eureka, Meyer and Lisbon.
- Farming method: Organic and inorganic methods affect the production by 10-30%. Organic methods that involve composting, no tilling and no chemical fertilizers can reduce the quantity produced by 20-30%. Thus, this input is also considered when generating the dataset.

For this framework, the following parameters are generated for each profile.

Table 3.3 Parameters for generation of prosumer profile

Parameters	**Value**
Postcodes	Victorian
Land size	474sm
Lemon varieties	Eureka, Lisbon and Meyer
Number of trees	1
Tree age	3-6 years
Lowest production	0-50 units
Highest production	1500 units
Harvest season	Winter or Summer
Per capita consumption	40
Family size	1-7
Farming method	Organic or Inorganic

- Lemon Consumption Rate: For prosumers, it is important to estimate their family consumption and calculate the surplus production that can be shared with the community or market. To do so, the per capita consumption of lemons is estimated and average family size is determined. Finally, prosumer consumption is calculated and averaged out to obtain the lowest production and highest production rates.
- Lemon Production Rate: As a lemon tree ages, its yield increases. When it reaches maturity after five years or so, it can produce an average of ~1500 lemons. The total amount produced also depends on whether organic or inorganic farming methods have been used. Therefore, the farming method used and the age of the lemon tree are combined to estimate the average production for a season or a whole year. Finally, the estimated average production amount is assessed and consumption is calculated to obtain the LYC and HYC. The LYC and HYC show the maximum contribution for the season that can be expected from a prosumer.

Variety	Farming-method	Season	Tree-age	Family size	Consumption	Postcode	LYC	HYC	longitude	latitude
Lemon-Lisbon	organic	June-Oct	4	6	240	3143	0	360	145.0194	-37.8589
Lemon-Eureka	organic	June-Aug	4	4	160	3055	80	440	144.9422	-37.7636
Lemon-Lisbon	inorganic	June-Oct	6	2	80	3143	520	1520	145.0194	-37.8589
Lemon-Eureka	inorganic	June-Aug	5	1	40	3004	460	1460	144.9702	-37.8442
Lemon-Eureka	inorganic	June-Aug	5	3	120	3053	380	1380	144.9661	-37.8036
Lemon-Meyer	inorganic	all year	5	4	160	3206	340	1340	144.9509	-37.8465
Lemon-Lisbon	inorganic	June-Oct	6	1	40	3141	560	1560	144.9913	-37.8407
Lemon-Meyer	inorganic	all year	4	5	200	3056	100	550	144.9601	-37.7663
Lemon-Lisbon	organic	June-Oct	3	4	160	3181	0	77.6	144.9955	-37.8547
Lemon-Eureka	inorganic	June-Aug	5	1	40	3121	460	1460	145.0018	-37.8233
Lemon-Eureka	organic	June-Aug	5	2	80	3056	320	1120	144.9601	-37.7663
Lemon-Eureka	inorganic	June-Aug	5	7	280	3182	220	1220	144.9795	-37.8653
Lemon-Meyer	organic	all year	5	4	160	3181	240	1040	144.9955	-37.8547
Lemon-Lisbon	inorganic	June-Oct	6	3	120	3141	480	1480	144.9913	-37.8407
Lemon-Lisbon	inorganic	June-Oct	3	7	280	3141	0	17	144.9913	-37.8407

Figure 3.7 Prosumer dataset

After determining the production-sharing rate, we randomly generate 300 production profiles (shown in Figure 3.7), which are then used to verify the proposed APCGs concept. Results, and discussion of the prototype is presented in the following section.

3.6.2 Results and discussion

In this section, the proposed framework's theoretical features are evaluated. As discussed in the theoretical fundamentals section (3.5), the framework development consists of two phases. Phase 1 involves the clustering of agro-prosumer profiles, and outlier detection; in phase 2, the agro-prosumer clusters are optimised, and the pre-conditions for groups are defined.

To start the verification process, phase 1 activities are carried out. Firstly, the agro-prosumer profiles are collected and the dataset is prepared and checked for data quality. For instance, the production and consumption of agro-prosumers are analysed and if the maximum production share is less than 50 for a season, that profile is discarded. For this framework, 300 prosumer profiles were obtained as a sample, of which five were discarded as their HYC was less than 50.

Next, the dataset consisting of prosumer profiles is partitioned according to suburb or municipal boundaries, and irrelevant profiles are removed. Kmeans techniques can be utilised to cluster groups or city's zone boundaries can be used for the purpose. Of the 300 prosumer profiles, 87 prosumers belong to "G-206 clusters" and 200 prosumers belong to "G-207 clusters". The remaining eight profiles are kept in a small extra cluster as outliers.

The resulting clusters, G-206 and G-207, are formed based on the geographical location and removing the outliers. These clusters are further partitioned into different prosumer groups based on their production rate using the hierarchical clustering method described in section 3.5. For G-206, hierarchical clustering resulted in four clusters. Figure 3.8 illustrates the number of prosumers allocated to G-206 clusters where c1, c2, c3 and c4 denote four cluster groups produced by the hierarchical method.

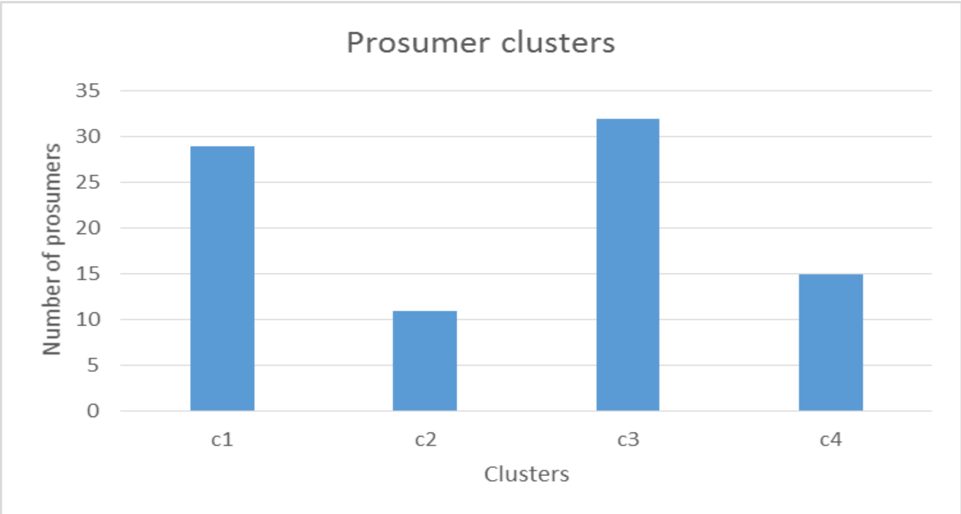

Figure 3.8 Number of prosumers in each of the four clusters

The same hierarchical clustering is done for the G-207 cluster, which resulted in eight clusters: c1 to c8 (Figure 3.9).

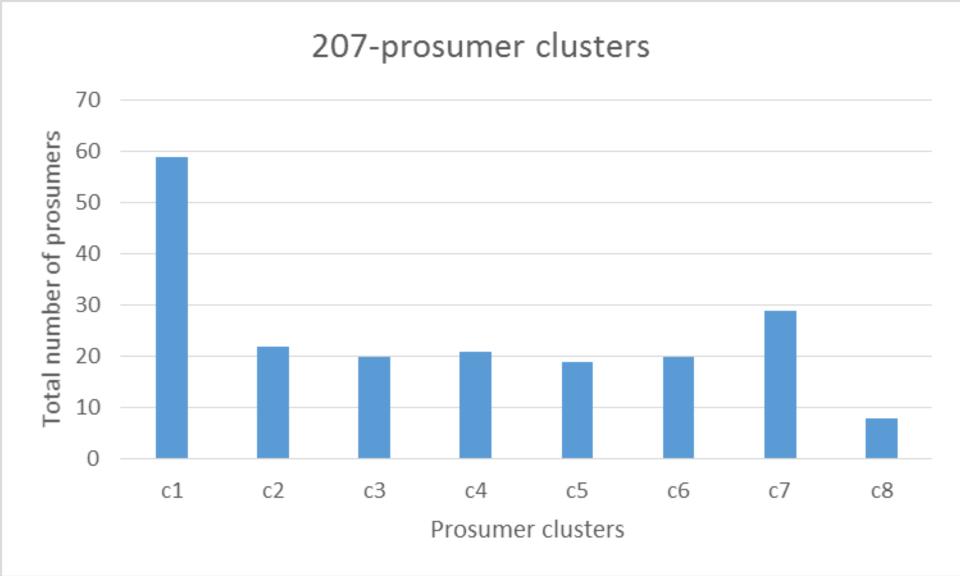

Figure 3.9 Number of prosumers in G-207 cluster

However, as shown in Figures 3.8 and 3.9, some clusters have a very large number of prosumers; for instance, there are more than 30 agro-prosumers in c3 of G-206, and nearly 60 in c1 of G-207. APCGs need to have a reasonable number of members in each cluster: small clusters can cause inefficiency or overheads, and large clusters can overproduce and cause storage problems or damage to the produce. Hence, in this case, the optimisation of the clusters is done by splitting the large clusters just to ensure an appropriate number of members.

In addition, Figures 3.8 and 3.9 show clusters which are too small where the number of agro-prosumers is less than or little more than ten. For example, cluster c2 in Figure 3.8 offers only 11 agro-prosumers and c8 in Figure 3.9 has only 8 agro-prosumers. If the APCG fails to supply an adequate amount of produce to the buyers or market, it might not enjoy good value or strong relationships in the long term and may become unsustainable. Therefore, in this scenario, consecutive prosumer clusters are merged to meet the amount of production required of members. The number of clusters are reduced by merging the neighbours into one cluster. These finalized clusters then constitute the APCGs.

The original clusters produced are optimised into most favourable number of groups (i.e. APCGs) by ensuring members count in each APCG fall within the range (i.e. between the defined minimum and maximum number of prosumers) and receiving the minimum amount of production from each APCG. For G-206, we divide the large clusters into two APCGs by splitting the production quantity further down (we assume 10 prosumers min. and 40 prosumers max.) in each APCG, and each APCG collectively produces quantity (at least---). These finalised clusters are illustrated below in Figure 3.10 for G-206 clusters.

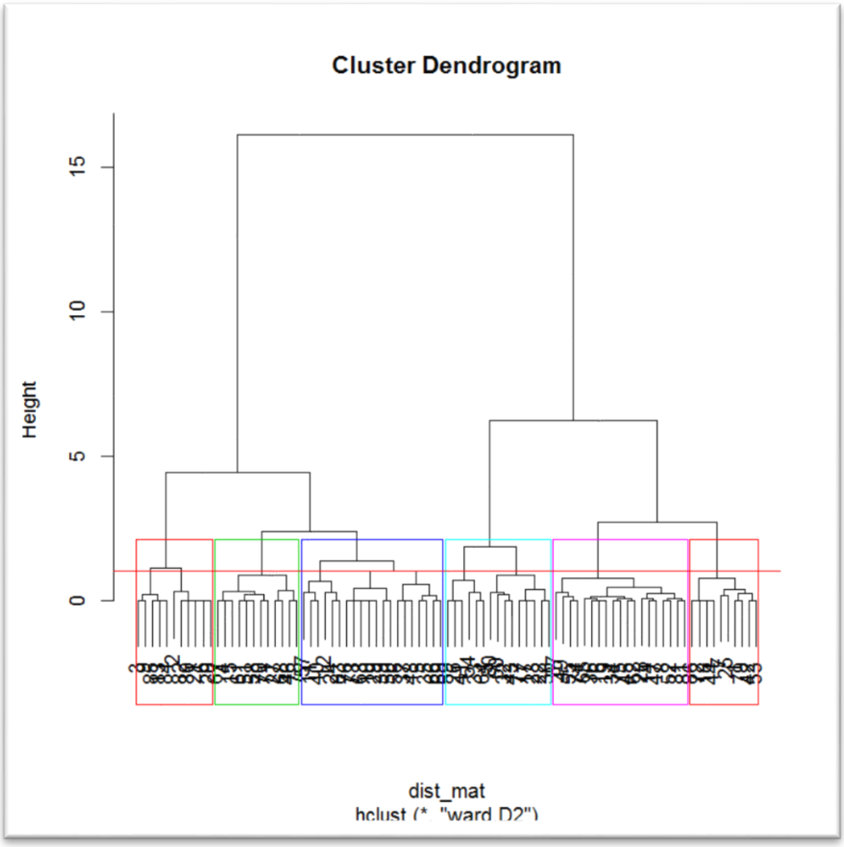

Figure 3.10. APCGs for G-206

Table 3.4 Hierarchical clusters for G-206

G-206				
Cluster	**Total number of Prosumers**	**LYC**	**HYC**	**Average production**
1	10	20	550	285
2	11	460	1560	1010
3	12	380	1420	900
4	20	220	1360	790
5	19	0	257	128.5
6	15	120	1000	560

Similarly, G-207 small clusters are merged into one by combining close production rates. The final clusters for G-207 are shown below in Figure 3.11.

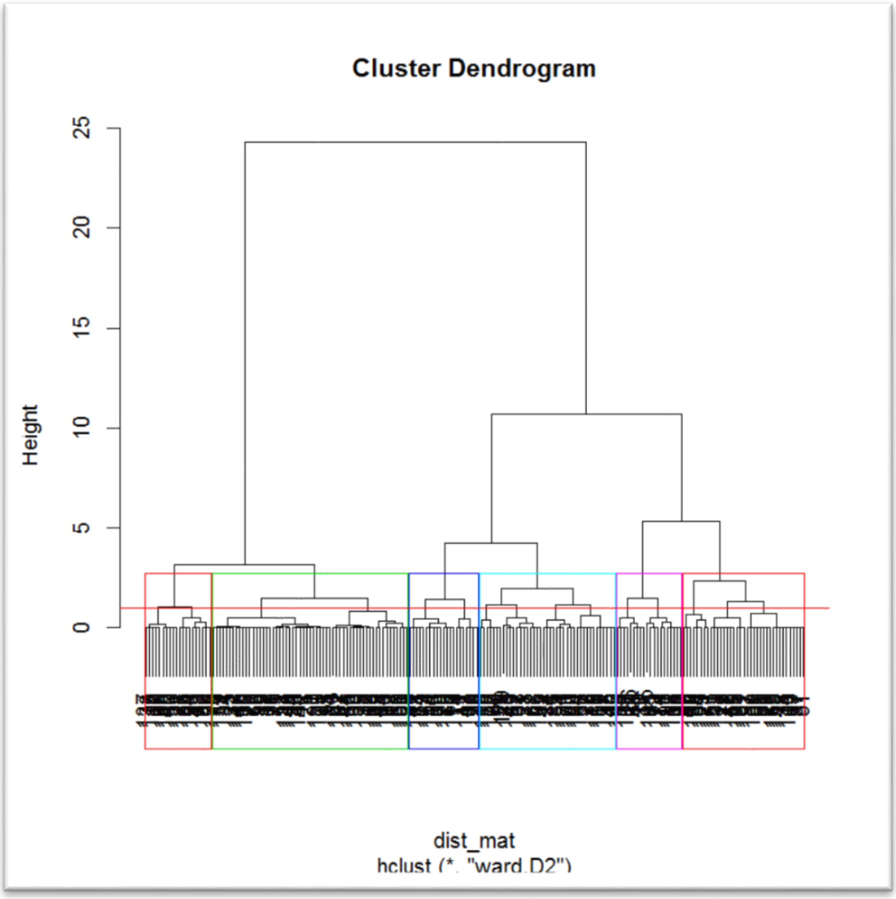

Figure 3.11 APCGs for G-207

Table 3.5 Hierarchical cluster output for G-207

G-207				
Cluster	**Total number of Prosumers**	**LYC**	**HYC**	**Average production**
1	59	0	257	128.5
2	41	320	1400	860
3	20	100	670	385
4	21	440	1560	1000
5	20	0	510	255

6	37	120	1260	690

Tables 3.4 and 3.5 illustrate the numerical distribution of prosumers into APCGs for G-206 and G-207 respectively. Using the distribution, similar patterns can be used to define and characterize the APCGs. Next, the pre-condition step is used to characterize the APCGs' entry requirements. Table 3.6 combines the average production and summarizes the pre-requirements for different APCGs during a season. The pre-requisites are provided to any interested prosumers to aware them with the entry requirements for APCG.

Table 3.6 Agro-prosumer community group pre-condition criteria

Agro-prosumer community groups	Total number of Prosumers	LYC	HYC	Average production
APCG1	59	0	257	128.5
APCG2	20	25	510	255
APCG3	20	100	670	385
APCG4	37	120	1260	690
APCG5	41	320	1400	860
APCG6	21	440	1560	1000

3.7 Conclusion

In this chapter, an APCG concept is proposed and its first framework which is to define agro-prosumers based on his/her interests, production behaviour, geographic location etc. Then categorize them into feasible APCGs for better structure and operability has been presented. The interested agro-prosumers are registered with their postcode, production details and personal interests. The pre-requisites for each APCG are determined which then can be referred when hiring new agro-prosumers. A prototype using R is presented in this chapter to verify the APCG working model. The next chapter will outline the new prosumer Engagement process.

Chapter 4

4 Agro-prosumer Engagement Framework

This chapter provides

- An introduction to a new agro-prosumer engagement process.
- A method to analyse agro-prosumers real time production behaviour for engagement.
- A conceptual process of outlining proposed framework requirements, design logic and theoretical fundamentals.
- Verification of framework using new prosumers dataset.

4.1 Introduction

In the last chapter, we established an agro-prosumer community groups (APCGs) concepts, where agro-prosumers' profiles are assessed and analysed to form different groups and derive each group's unique pre-requisites. The framework utilizes agro-prosumers' production history as a prerequisite of APCGs. One can join APCGs while checking perquisites, however, it is important to evaluate the agro-prosumers' dynamic production profiles before hiring them permanently into APCG network. In addition, rather than relying on historic production behaviour, it is important to use dynamic production profiles, which will give a better understanding of the prosumer's commitment to supporting the APCG.

In this chapter, we propose a method for recruiting new agro-prosumers for an APCG. The engagement is based on an evaluation of their dynamic commitment conducted over a defined period.

In this chapter, section 4.2 gives an outline of the agro-prosumer engagement framework. In section 4.3, the requirements for this framework is discussed, and section 4.4 presents the design logic. The theoretical fundamentals of the framework is discussed in section 4.5, the verification is shown in section 4.6, and section 4.7 concludes the chapter.

4.2 Outline

The design proposed for the engagement of a new agro-prosumer into APCG is a latest approach. This section provides an outline of the design framework as shown in Figure 4.1.

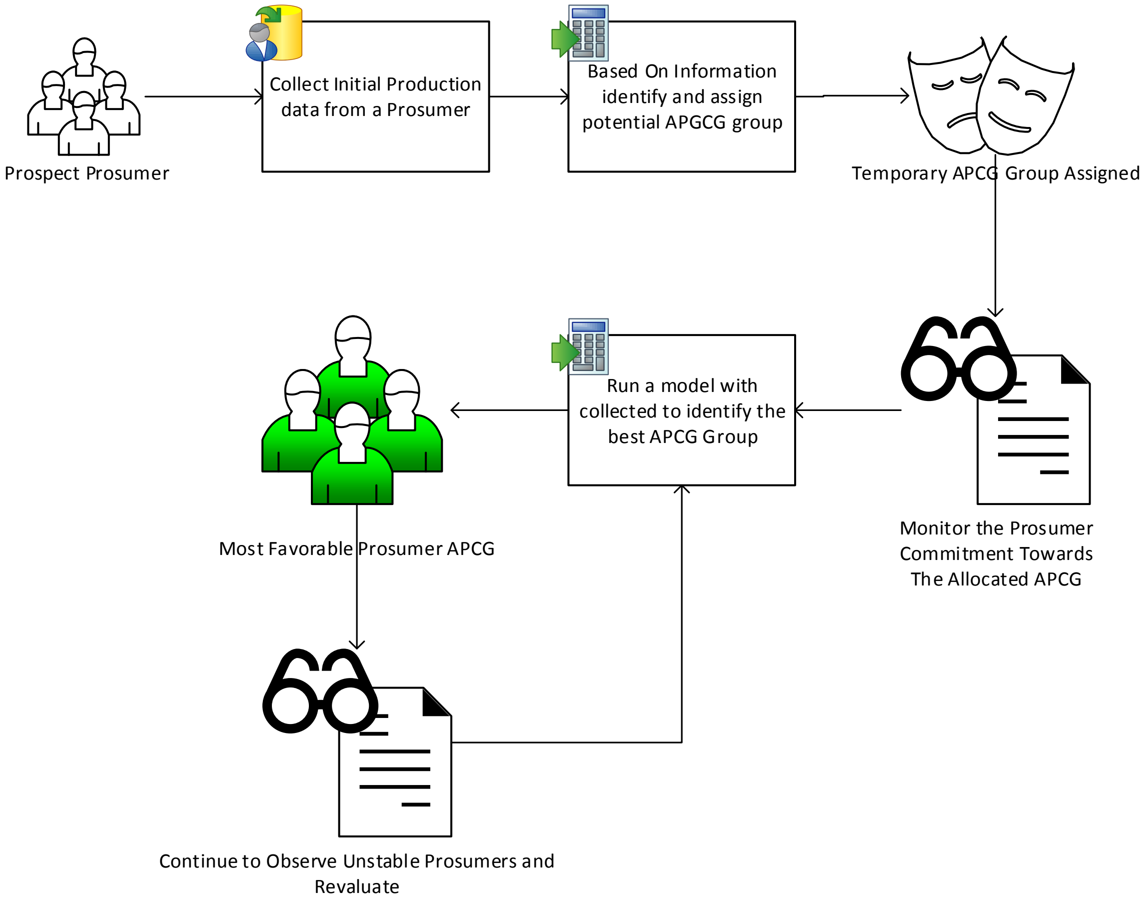

Figure 4.1 Outline of agro-prosumer Engagement framework

New agro-prosumers who are interested in joining the APCGs, and their dynamic behaviour profiles, are collected as input for this framework. These agro-prosumers are called as “enrolled agro-prosumers” who are assumed to be new to the APCGs; thus, because there is lack of experience or historical production data, dynamic production needs to be determined. The final outcome of this framework is the recruitment of prospect agro-prosumer to suitable APCGs. A notion “permanent APCG” is used to represent the finalized APCG of the prosumers.

Figure 4.2 represents the communication points happen between the new agro-prosumer and the web module. Agro-prosumers register their interest towards the APCG, and get notified about different APCGs pre-requisites and activities required to be delivered in order to become a member of APCG. The two main requirement of an APCG as discussed in the previous chapter are Lower Yield Count (LYC) and Higher Yield Count (HYC), where the LYC of an APCG is the lowest amount of produce expected from a new agro-prosumer to share in order to be accepted as a member of an appropriate APCG, and HYC is the highest amount of produce expected that the new agro-prosumer may offer to share as a member of an APCG.

The agro-prosumer chooses the APCG of most interest to him/her, termed “chosen prosumer community group”. The next step is the evaluation period which is denoted by an algorithm with inputs such as: type of farming and fertilizers, consumption pattern and variety of vegetation (lemon variety) for a certain number of seasons. This evaluation is undertaken prior to assigning a prosumer to a particular APCG. Accordingly, a set of rules must govern the evaluation period for prospect agro-prosumer. This evaluation period plays a critical role in providing a better understanding of the agro-prosumer’s production behaviour, rather than relying only on the amount of produce. This is a necessary step before a prosumer is accepted as a permanent member of an APCG.

The evaluation involves several important tasks: depending on the lemon variety, for two consecutive seasons, agro-prosumers are asked to report their consumption pattern, and production is predicted using an algorithm and other inputs such as farming method (organic or inorganic) age of trees, total number of trees, variety of lemon, and fertilizers used. The algorithm for dynamic production is illustrated by means of a chart in Figure 4.2.

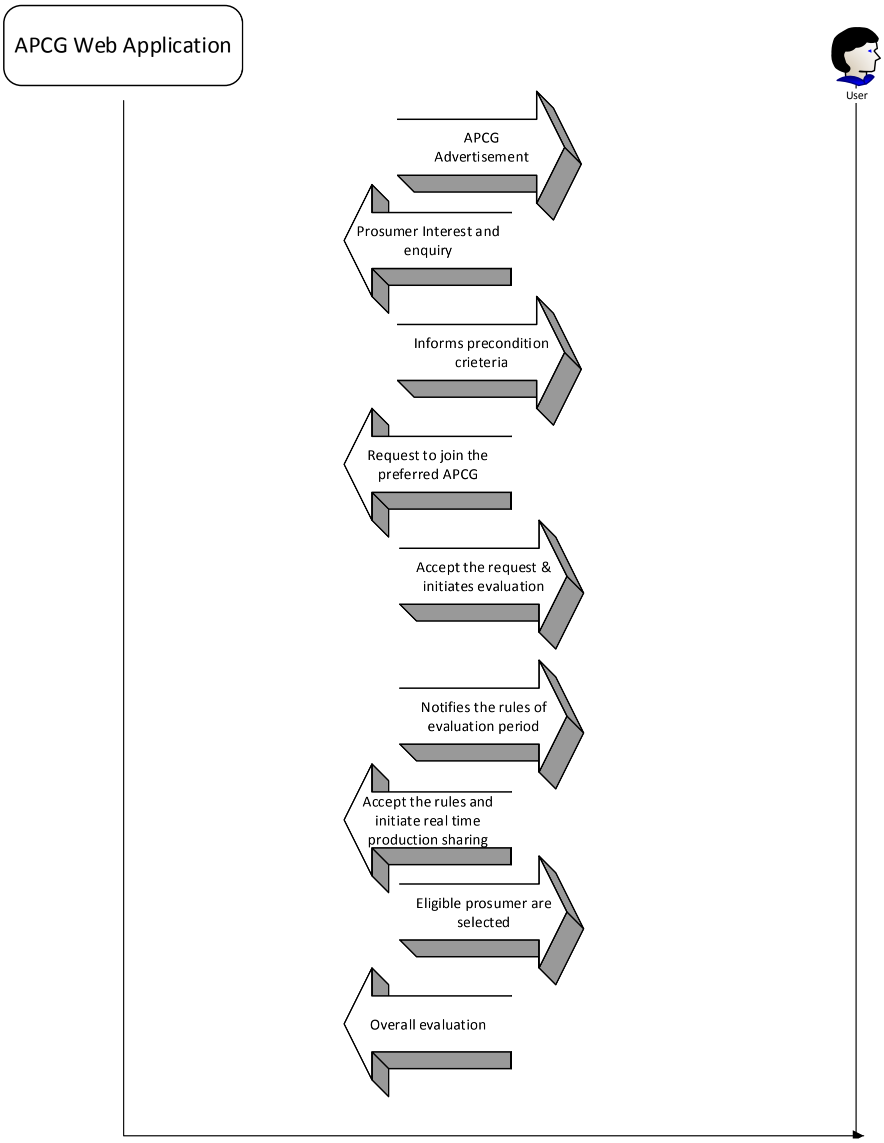

Figure 4.2 Abstract flow of a new agro-prosumer evaluation

Using the algorithm above, the amount of dynamic production is calculated and the prosumers' production performances are assessed for each season. If these performances meet the lowest commitment expected by the APCG, then prosumers become "eligible prosumers". This lowest commitment expected by an APCG is termed "eligible energy threshold". If agro-prosumers' performances meet the pre-requirements established by their preferred APCGs, they are transferred to the corresponding APCG on a temporary basis. However, if an agro-prosumer's performance does not satisfy the production requirements of his/her chosen APCG, but able to meet those of other APCGs, he/she will be advised to join the latter. Initially, prosumers are assigned to a "temporary APCG" during the evaluation period. Closer to end of the evaluation period, prospect agro-prosumers' overall performance is evaluated and they are categorised as either "reliable prosumers" or "unreliable prosumers" based on the level of satisfying pre-requirements of APCG. The “reliable prosumers” are those who meet the pre-requirements of any APCG and have achieved successful results, besides the “unreliable prosumers” are those who were unable to meet the entry requirements of any APCG during the evaluation period. The “reliable prosumers” are then accepted to the most appropriate or closest APCG as permanent members. This engagement of new prosumer is called as “community group membership”. Later on, the prosumer is recruited to the APCG and begins the “production contract” with the APCG and converted to a permanent community group membership for it. The production contract is a smart contract (for blockchain platform, chapter 6) between the prosumer and APCG platform which starts when the prosumer is recruited to a specific APCG. The information stored in the smart contract outlines the rules and conditions governing the relationship between prosumer and APCG. Following section discusses the requirement for the solution framework.

4.3 Requirements

Next step of a conceptual process is discussing the requirements of the proposed solution. Following are the requirements for new prosumer engagement process:

- **A technique to evaluate prosumer:** In order to evaluate a new agro-prosumer, the evaluation period and inputs must be clearly defined; for instance, the duration of time or seasons for the evaluation period must be defined, together with inputs for production estimates and their magnitude to evaluate the prosumer’s commitment and fulfilment of the APCG’s expectations.
- **Approach to evaluate and monitor agro-prosumers real time production behaviour:** Agro-prosumer must be evaluated over a specific time period with a focus on analysing real time production behaviour before recruiting them to APCGs. Also it will help in recruiting new members who doesn’t have any background. For instance, if an agro-prosumer is completely new to farming, an ongoing evaluation process is essential because there is no historical data available to predict his/her behaviour for the purpose of allocating the prosumer to an appropriate APCG.
- **Determining reliable prosumers prior to APCG allocation**: The proposed solution should identify irregular production behaviour and clearly differentiate the reliable prosumers from the others during the process. If a prosumer is new to the production process, it makes the process complicated and the prosumer may require guidance and support for his/her farming endeavours.

This concludes the first stage of conceptual framework. The design logic discussed next shows the design decisions for each of the aforementioned requirements, together with a detailed discussion.

4.4 Design Logic

The design logic for all the requirements discussed in the previous section must be presented as a next step. This summarises the second stage of the conceptual process. Figure 4.3 gives an outline of the design decisions.

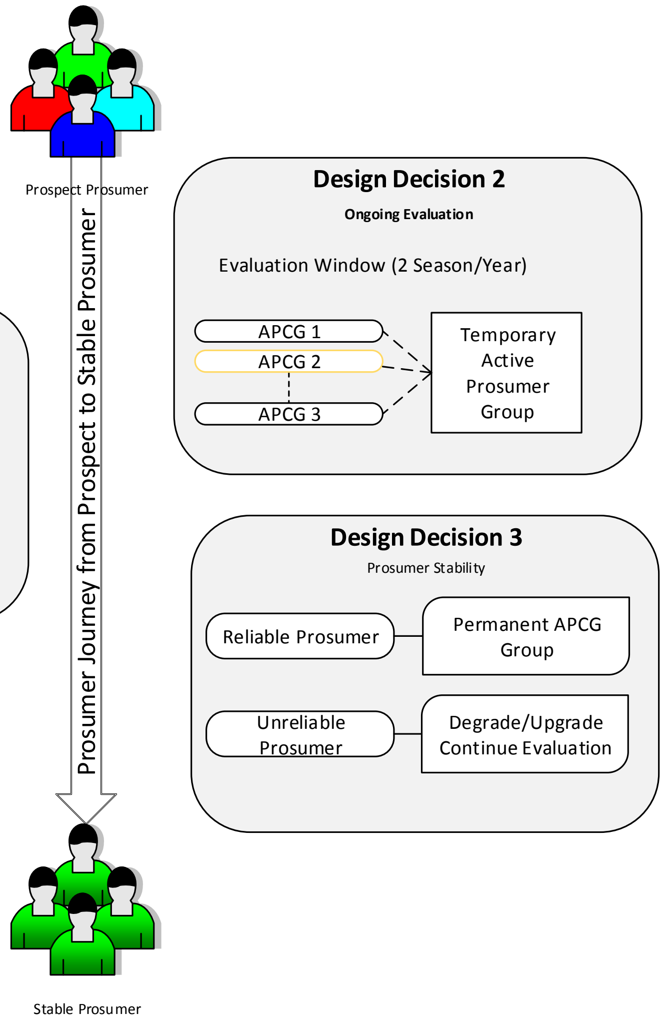

Figure 4.3 Design Outline of framework proposed

Design logic 1

For new agro-prosumers' recruitment, it is important to analyse their production contribution against their preferred APCGs requirement. Therefore, the evaluation technique is the fundamentals of the agro-prosumer engagement process. A performance scale is proposed to evaluate the agro-prosumers' capability of meeting the pre-requisites. A technique proposed here is an approach used to engage the agro-prosumer to the most appropriate APCG. This design logic cover up the requirement of determining an approach to calculate the production-sharing behaviour.

Design logic 2

Two years containing 2-4 seasons, based on the variety of lemons, is established as the "evaluation period" for the new prosumer. This step satisfies the requirement of continuously evaluating and monitoring an agro-prosumer before offering membership of a specific AOCG.

Design logic 3

In order to distinguish reliable from unreliable agro-prosumers among the prospect agro-prosumers, the stability index (SI) is used to analyse the agro-prosumers' overall capability post evaluation period. This method confirm the requirement of identifying reliable and unreliable agro-prosumers.

These design decisions should be considered when developing an agro-prosumer engagement framework. In the next section, the theoretical fundamentals for the proposed solution is explained.

4.5 Theoretical fundamentals

Next step for agro-prosumer engagement framework is building the theoretical base. This is the third stage of the conceptual process. This stage is further divided into four components, which are explained below.

The framework has four components

- (i) An approach to evaluate agro-prosumers' production performance;
- (ii) Agro-prosumers' transaction assessment during the evaluation period
- (iii) An approach to analyse agro-prosumers' stability; and
- (iv) Agro-prosumers recruitment to a specific APCG after the evaluation period.

The varying nature of agro-prosumers' production behaviour is evaluated using the above approaches, and allocates them to a temporary "variable APCG". Later on, the prosumers' overall behaviour is stored and evaluated prior to recruitment to a specific APCG, i.e., to one of the final APCGs. As discussed in section 4.3, the requirements for the proposed solution are covered by four components listed above. These four components are discussed in detail below.

4.5.1 Approach to evaluate prosumers' production performance

Finding an approach to estimate agro-prosumers' performance is the first component of the evaluation technique, which helps in understanding the evaluation period activities and evaluation inputs.

4.5.1.1 Agro-prosumer evaluation measures

As discussed in previous chapters, the "evaluation period" is an established period of consecutive seasons during which the production behaviour of new agro-prosumers who are interested in joining an APCG, is evaluated. The evaluation period is divided into two seasons per year in Australia: winter (i.e., March-August) and summer (i.e., September- February). These winter and summer seasons show non-overlapping, mutually exclusive time periods and are assigned with a production transaction between agro-prosumer and the APCG module using production value.

Agro-prosumers' production data such as family size, farming methods, lemon variety and number of trees and their respective ages, are collected as input to evaluate their consumption pattern and production performance for two season or annually. Agro-prosumers' surplus production is considered as the final value for one season/year. Thus, prosumers' performance is estimated using that final value, and is evaluated for each season. Next section explains the approach used to determine the prosumers' performance for each season during the evaluation period.

4.5.1.2 The proposed method

This approach requires two inputs; the input from the agro-prosumer and the input from the APCG module as shown in Figure 4. Inputs from the agro-prosumer include production summary for a season and the prosumer's preferred APCG. The APCG module's input comprises the pre-condition criteria of the available APCGs.

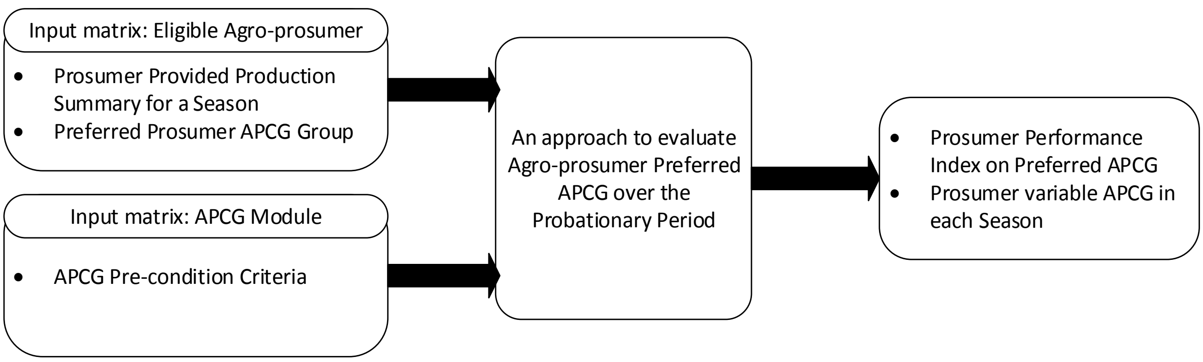

Figure 4.4 An outline of an agro-prosumer evaluation approach

A probabilistic approach is designed here to evaluate agro-prosumers' production performance that results in a performance scale formation. Performance indices are used to anticipate the capability of an agro-prosumer in meeting the pre-requisites of his/her preferred APCGs. Different levels of success and failure are represented using a four-point scale as shown in Table 4.1. In table 4.1, each performance index shows different value or success rate of performance in the production behaviour.

Table 4.1 Performance scale interpretation

Performance Index Interpretation	**Success/Failure Rate**	**Performance Scale**
Complete Success	100%	3
Intermediate	90-99%	2
Entry	80-89%	1
Failure	0-79%	0

The performance scale ranges from 0 to 3, where 3 represents the complete success or match, and the minimum success rate is 80% for meeting the pre-condition criteria. If the success rate is less than 79%, it will be considered as a "failure".

The performance scale used here has single-integer values as it is difficult to use extreme values like only high or low, to measure prosumer behaviour. Prosumer production behaviour can vary and to accurately measure it, the various levels of performance must be identified. A performance scale with a value from 0 to 3 will help to indicate prosumers' performance for APCGs development.

- Complete success: The highest point on the performance Scale is 3, which indicates "Complete success". This scale suggests 100% success rate in interacting with the prosumers' production-sharing process. This level of performance according to the PS in table 4.1 suggests that the prosumer is strongly suited to his preferred APCG and meets the desired pre-condition criteria.
- Intermediate success: This level denotes 90-99% of success rate in interacting with prosumers' production behaviour. Performance Scale 2 shows that it is the "medium

success" level. This scale value suggests that in meeting the prosumers' preferred APCG requirements, prosumers' performance reliability is good.

- Entry success: Performance scale 1 indicates "Entry success". This scale value suggests 80-89% success rate while satisfying the pre-requirements of the preferred APCG's. This performance index scale suggests that the prosumer is slightly reliable in meeting the desired pre-condition criteria of his/her preferred APCG.
- Failure: 0 reflects the lowest scale in performance, indicating "failure". This level depicts 0-79% rate of success in fulfilling the pre-requirements. Thus, this level shows that the prosumer's performance is not reliable enough to meet the pre-condition criteria for the APCG. Hence, the prosumer with this index could be matched with other APCG rather than the preferred one.

The mathematical expression of performance indices is given in equation 4.1. Let's say for a season *j*, prosumers' success rate P_{ij} being allocated to chosen temporary APCG (C_p):

$$Rate(P_{ij} \in C_p) \begin{Bmatrix} 100\%: if\ E_{ij} \geq L_p \\ \frac{E_{ij}}{L_p}: if E_{ij} < L_p \end{Bmatrix}$$ Equation 4.1

Where,

$Rate(P_{ij} \in C_p)$	Performance Score	Interpretation
100%	3	Complete success
90%-99%	2	Medium success
80%-89%	1	Entry success
<79%	0	Failure

P_{ij} is an i^{th} agro-prosumer's performance in the j^{th} season, C_p is the chosen APCG, E_{ij} is the real time production commitment of i^{th} agro-prosumer and L_p is the production threshold of agro-prosumers preferred APCG. Now suppose if group A has 500 threshold (Lp) and agro-prosumer Ap was able to produce 450 (Eij), then the performance score ($Rate(P_{ij} \in C_p)$) will be 450/500 i.e. 90%. Thus Ap's score would be rated as 2.

4.5.2 Agro-prosumers' transaction assessment during the evaluation period

For ongoing assessment during the evaluation period, agro-prosumer is aimed to assign into his preferred APCG for each season. The assessment process is shown in Figure 4.5. The key steps of the process are as follows: the prospect prosumer is asked to submit records of production in real time for "n" seasons during the evaluation period. For each season, dynamic production amount is compared with the minimum threshold (E_{th}), which is the minimum requirement of any APCG. If the prosumers' production is equal or greater than the E_{th}, the prosumer is viewed to be an eligible prosumer.

Next, if a prospect agro-prosumer receives 'eligible prosumer' status during his/her first season, she/he will be promoted to the next season and then to following seasons. However, if she/he fails to meet the eligible agro-prosumer requirement, in the first production season, the evaluation period will be extended with more seasons.

However, if the new agro-prosumer is not able to match the minimum threshold (E_{th}), then the prosumer's evaluation period is extended by another season and the prosumer remains under evaluation until succeeded. On the completion of the evaluation period, prospect agro-prosumers' stability will be analysed using stability index, which is discussed in next step.

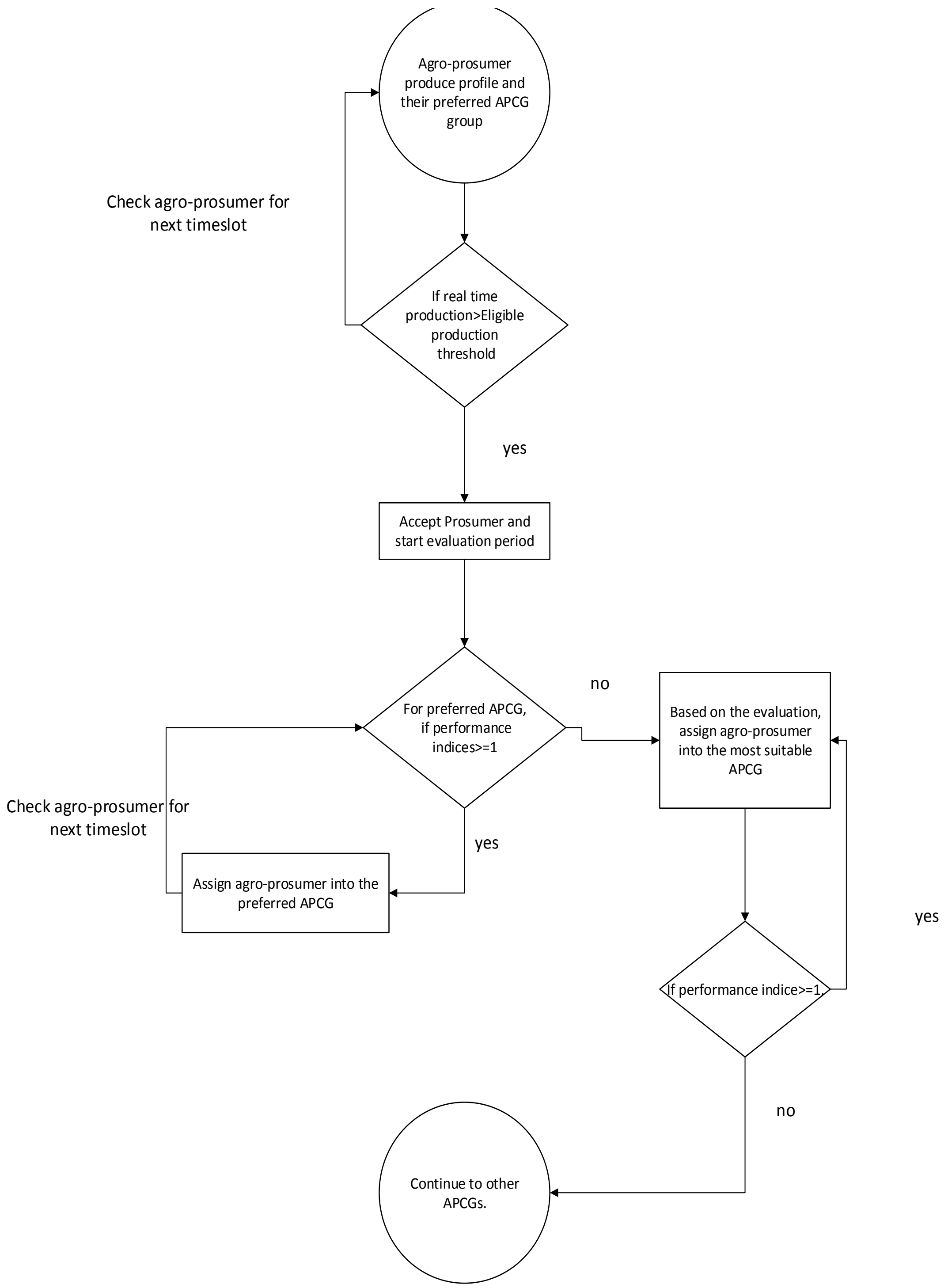

Figure 4.5 Assessment of agro-prosumer evaluation process

4.6 An approach developed to analyse agro-prosumer stability

The stability of an agro-prosumers' reliability is estimated for his/her preferred APCG, as well as for those assigned throughout the evaluation period. Figure 4.6 shows a process to obtain prosumers' stability for agro-prosumers' chosen APCG.

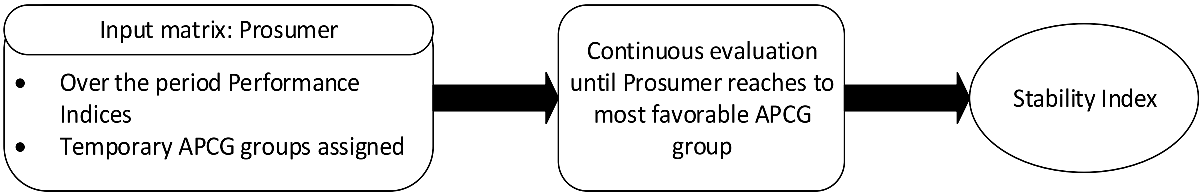

Figure 4.6 Approach to determine stability index for new prosumer in APCG

During evaluation period, for each season, agro-prosumers' performance index values are taken as an input along with their temporary APCGs. Equations 4.2 and 4.3 formulates a mathematical equation for the approach. SI represents the stability index which is used to determine the feasibility that prosumers will remain in their preferred APCG. The output for I index is between 0 and 3, and a higher I shows high chances of prosumers remaining in their preferred APCG:

$$I_{pi} = \frac{\sum_{j=1}^{ns} PX_{ij}}{ns} \qquad \text{Equation 4.2}$$

Above I is the stability index of the i^{th} prosumer with respect to chosen APCG (C_p), PX_{ij} is an i^{th} prosumers' performance index in the j^{th} season and ns is the number of seasons where the prosumer is assigned to his/her chosen APCG. To determine most suitable APCG for an agro-prosumer, rate of engagement to a specific APCG is calculated using equation 4.3. For example, if the agro-prosumers' rate is higher for APCG1 than other APCGs, than the chosen APCG1 is seen as the most favourable APCG for that prosumer's engagement.

$$Rate(P_i \in APCG_{Fr}) = \frac{count\ of(APCG_{Fr})}{ns} \qquad \text{Equation 4.3}$$

Where P_i is the i^{th} prosumer, $APCG_{Fr}$ is the r^{th} temporary APCG, count of ($APCG_{Fr}$) shows the total number of times the prosumer is selected to r^{th} temporary APCG during the evaluation period and ns is the number of seasons.

The next section discusses agro-prosumer engagement to the permanent APCG based on the previously-described method.

4.6.1 Agro-prosumers engagement to the permanent APCG after the valuation period.

Agro-prosumer engagement to the most suitable APCG is analysed in this step. The overall performance of prospect agro-prosumers overall performance is assessed at the end of the evaluation period. Figure 4.7 is a flowchart showing this process.

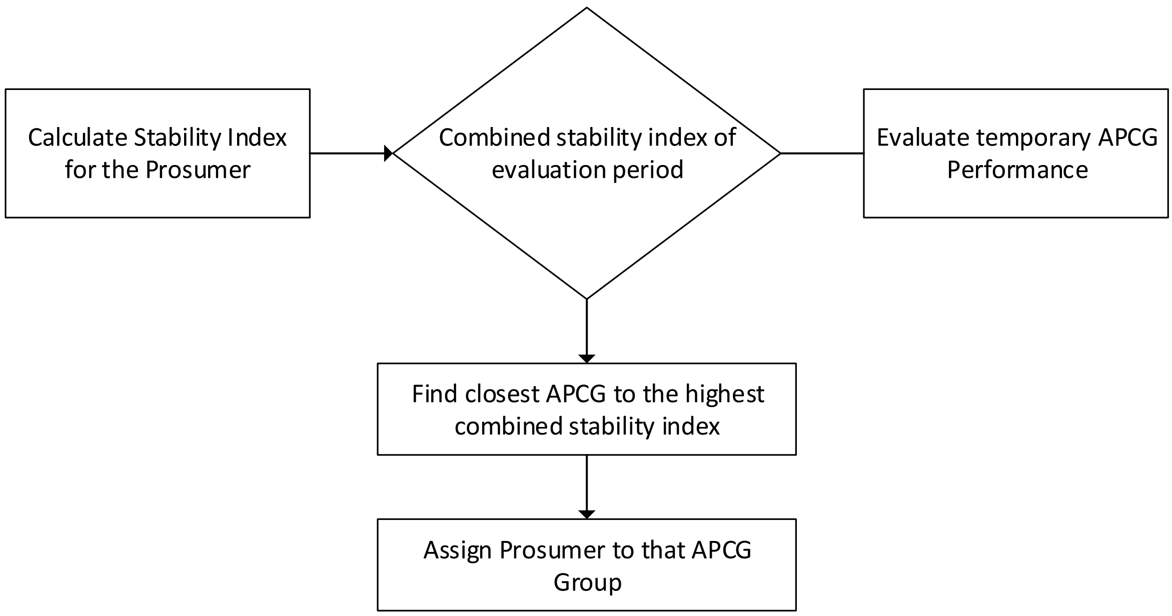

Figure 4.7 Agro-prosumer Engagement process

As discussed in the previous section, the Stability Index, based on an agro-prosumer's performance index, is calculated throughout the evaluation period. Additionally, agro-prosumers' rate of staying in temporary APCGs during the entire evaluation period is assessed. Equation 4.4 is utilised to identify the combined value of the agro-prosumer being allocated to the permanent APCG. The APCG which shows the highest joined index is chosen as that prosumer's final permanent APCG.

$$IPr(P_i \in APCG_{Fr}) = I_{pi} \times Rate(P_i \in APCG_{Fr}) \qquad \text{Equation 4.4}$$

Next section is about verifying the concepts discussed for the prosumer Engagement framework.

4.7 Verification of the agro-prosumer engagement framework

For verification and validation of the agro-prosumer engagement framework, the solution framework is simulated using MATLAB and Excel. MATLAB is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages (Wikipedia).

Firstly, the experimental setting for the proposed framework is discussed.

4.7.1 Experimental Setting

The setting here is a basic set-up for the examination of the proposed framework. To verify the proposed algorithm, 50 agro-prosumers production profiles were generated, assuming that these 50 agro-prosumers have shown interest in joining APCGs. For dataset generation, production behaviour along with consumption patterns from chapter 3 are used. Data is obtained for summer and winter seasons for four APCGs that are defined and characterized in chapter 3. Four seasons are used for the evaluation period: two summers and two winters. Thus, a prosumer is evaluated over a two-year period.

The prototype parameters are listed below in Table 4.2.

Table 4.2 Prototype parameters

Prototype parameters	Value
Eligible production threshold(average)	25
Prospect prosumers	50
Evaluation period	2years
APCG1	0-250
APCG2	25-550
APCG3	100-670
APCG4	120-1260

Results are presented in the next subsection using the above parameters.

4.7.2 Observation, Results and Discussion

The previous section discussed the important role played by evaluation techniques in the engagement framework. Prototypes are conducted to verify these techniques and help to verify the agro-prosumers' production-sharing behaviour, which is determined during the evaluation period.

Eligible agro-prosumers are identified during the evaluation conducted after each season of the evaluation period. Only those agro-prosumers who satisfy the "eligible production threshold" in the first season can proceed to the next season. Additionally, eligible agro-prosumers choose their preferred APCG, and cannot change the chosen APCG until the evaluation period is finished; thus, the chosen APCG remains fixed for four seasons.

However, the eligible agro-prosumers may show irregularity while satisfying the chosen APCG's pre-requirements during the evaluation period. To solve this issue, we already mentioned in section 4.5 that the prospect agro-prosumer is required to meet the lower threshold value of the preferred APCG to be able to meet the evaluation criteria. Additionally, to determine the extent to which a prospect agro-prosumer meets the pre-requirements of the chosen APCG, four performance indicator groups are introduced with values: "3", "2", "1" and "0" indicating "complete success", "intermediate success", "entry success" and "failure", respectively.

In this prototype, the prospect prosumers' capability in meeting their chosen APCG's pre-requirements are assessed at first. Figures 4.8, 4.9 and 4.10 show the percentage of prosumer allocation to different performance indicator values over the evaluation period i.e. four seasons for different APCGs (APCG1, APCG2, and APCG3).

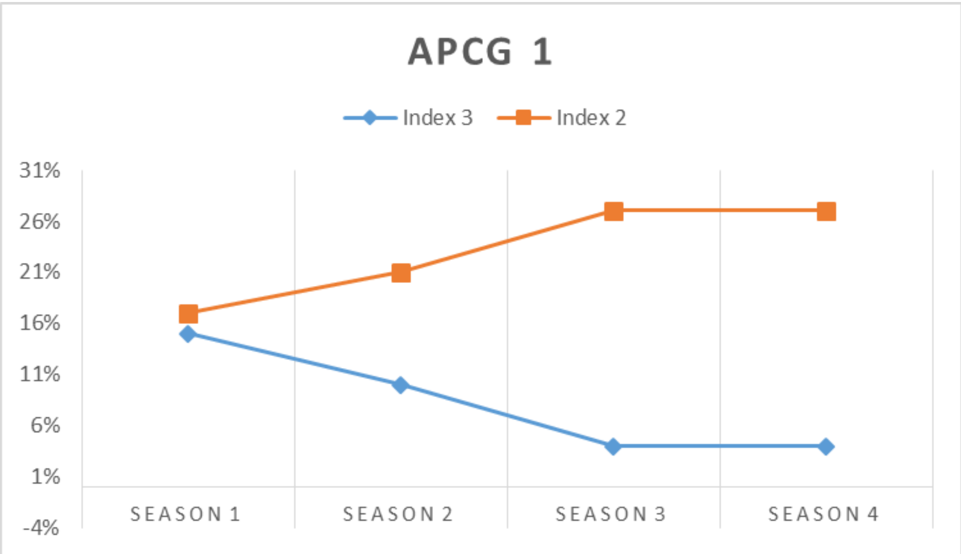

Figure 4.8 APCG1 agro-prosumer engagement pattern

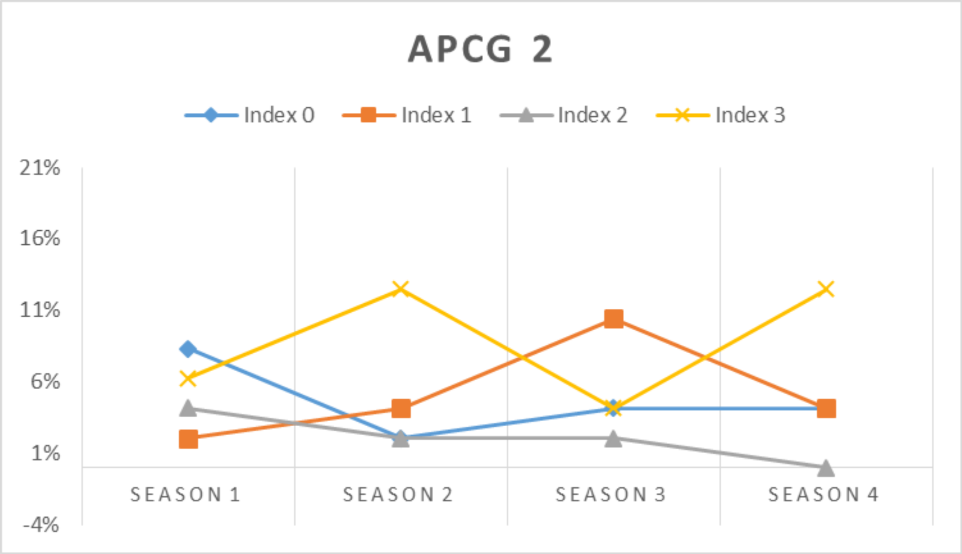

Figure 4.9 APCG2 agro-prosumer engagement pattern

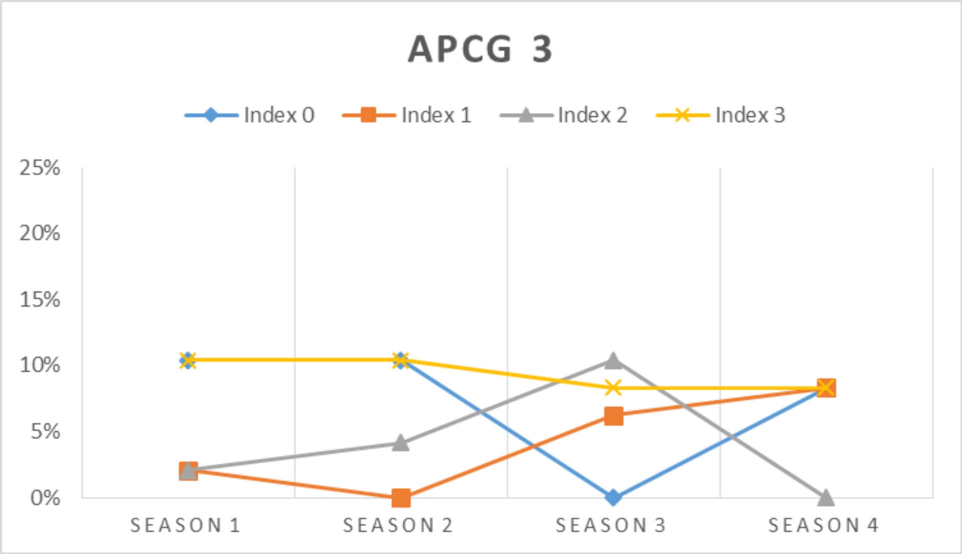

Figure 4.10 APCG3 agro-prosumer engagement pattern

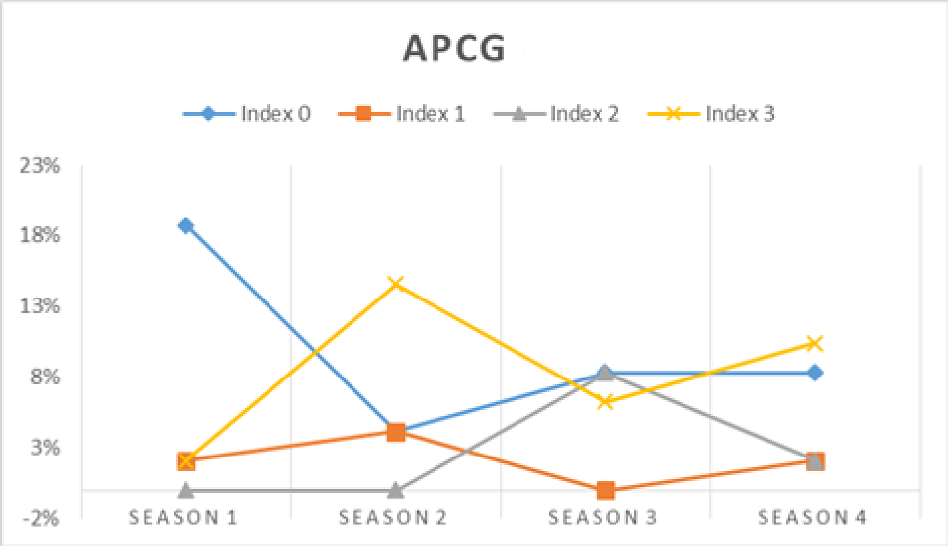

Figure 4.11 Combined APCG engagement pattern

Hence, when the agro-prosumer selects APCG1 as his preferred APCG, there is a strong chance that he/she will retain performance scale 3, i.e., “complete success”, throughout. This is because APCG1 has the least upper and lower thresholds for production value. On the other hand, when the prosumer chooses APCG4 which has more stringent pre-requirements, there is a strong chance that the agro-prosumer will retain the performance scale “0”, and will be offered membership of an APCG not of his/her choosing.

Moreover, agro-prosumers’ performance indices for each season and the stability indices are combined and used to allocate him/her to a specific APCG. Figure 4.11 illustrates the combined indices for each APCG. The X axis represents the prosumer identifier with respect to a particular APCG.

Finally, the agro-prosumers are allocated to the APCG based on their combined index. If the combined index is higher for the preferred APCG than that to which they have been allocated, they will be allocated to the most suitable APCG.

4.8 Conclusion

This chapter presents an agro-prosumer engagement framework which is a new approach of recruiting agro-prosumers in agricultural sector. In this framework, an agro-prosumer is assessed throughout the evaluation period, to check his feasibility of meeting the APCG’s pre-requirement and his stability is estimated. Finally agro-prosumer is engaged or recruited in the most appropriate APCG. Prototype results are used to verify the proposed framework. This chapter contributes to the APCG framework by building the second component of APCG framework.

After new prosumers engagement in the APCGs, another key challenge is to make APCG a goal oriented network. Identifying goals and managing diverse ones require a framework for a community platform. Chapter 5 describes in detail the framework for APCGs multiple goal management.

Chapter 5

5 Goal Management Framework

This chapter provides

- An introduction to a goal oriented APCG.
- A framework to manage diverse goals and resolve conflicting goals to achieve ideal set of goals for APCG.
- A conceptual process of outlining proposed framework requirements, design logic and theoretical fundamentals.
- Verification of framework using new prosumers dataset.

5.1 Introduction

In chapter 3 and 4, we presented two frameworks: one defining APCG concept and outlining categorization of different groups and other for the engagement of new agro-prosumers to specific APCGs. These frameworks underpin the construction of a community-based production network i.e. APCG.

After initiating the concept of APCGs, one of the critical requirement is to make APCGs goal-oriented. This can be done by determining the overall community objectives of APCG and, subsequently, finding an ideal set of goals by effectively managing conflicts among the multiple goals of the various APCG groups.

Goal management can be challenging in a community-based network due to multiple conflicting issues such as production and carbon sequestration goals. Several situations can occur where one objective is achieved but at the expense of another. For instance, in order to improve carbon sequestration in soil by APCGs, organic/ecological farming ways must be practised; however, organic farming methods yield less, which in turn affects the collective produce quantity of the APCGs and subsequent income. Thus carbon sequestration objective is inversely related to income objective or production quantity. However production quantity is directly related to income. Therefore, it is a requirement that multiple goals are compromised with respect to the limitations. Additionally, it is important to understand the shift in requirement specification in order to achieve a satisfactory level of the goals.

These factors demands an effective framework for the management of multiple goals. In this chapter, an effective framework to define and create an optimised set of overall goals is developed. The MCGP technique is used as the goal programming method to effectively negotiate among the multiple goals. Section 5.2 explains the features of the framework. Followed by section 5.3, where the requirements of the framework is discussed; the design logic is presented in section 5.4, followed by the theoretical fundamentals in section 5.5. The verification of the framework is conducted in section 5.6, and section 5.7 concludes the chapter.

5.2 Outline of the Goal Management Framework

Figure 5.1 shows the outline of the goal management framework.

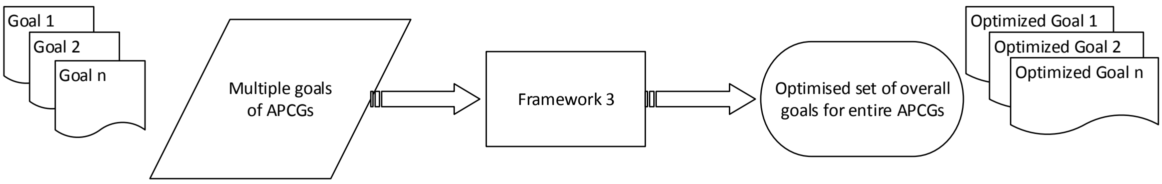

Figure 5. 1 Concise Outline of the goal management framework

In the framework, multiple goals for agro-prosumer community groups are taken as an input. As discussed, the main purpose of creating APCG is to support environmental sustainability, therefore, objective such as maximizing carbon capture is top priority. A number of other motivations can become the goals of APCG such as satisfying food security within the network, earning income and incentives, forming stable network, reducing maintenance cost etc. In addition APCG can also be seen as a solution to various existing issues such as food traceability, transparent carbon-transaction, production sharing network, source of income etc. which apparently comes with the APCG platform. To build the solution framework following objectives have been considered as the multiple goals of APCGs

(i) Maximum carbon capture objective. This objective relates to capturing the maximum amount of carbon dioxide while growing vegetables and fruits for APCGs' production sharing.

(ii) Food security within the network. The aim here is to satisfy the local prosumers' vegetable/fruit consumption demand within each APCG.

(iii) Providing local food access to wider community. The aim here is to fulfil the locally grown vegetables/fruits demand by selling APCG's production with local (suburb) buyers or grocery stores/supermarkets etc.

(iv) Income and Incentive objective. The goal is to achieve a certain income or profit by selling surplus crops or carbon tokens.

(v) Maintenance cost reduction objective. This aims to reduce the cost involved in harvest operations, distribution process and maintenance activities, and

(vi) APCG stability objective. This aim is to encourage the local prosumers to participate in the production/carbon sharing community network.

In the goal management process, the objectives of prosumer community groups are identified as single set of goals, and attempts are evaluated to achieve the most favourable goals. Once the overall optimised set of goals is obtained, these goals can be divided among the APCGs specific goals (specific goals is not a scope of this thesis). Following section, discusses the requirements for this research.

5.3 Requirements

Requirements for the proposed framework discussed in this section. This is the first stage of the conceptual process in which the requirements are stated and prioritized. The following are the requirements of the solution:

1. **Multiple conflicting goals are classified within APCG:** Multiple goals within an APCG are clearly outlined in the initial step. Here, all the essential conditions along with the desired objectives that are not among the must have goals must be clearly defined.
2. **Goal evaluation and prioritization process:** The evaluation of the goals, and their priorities based on their relative importance, are the next step of goal management system.
3. **Coming up with a satisfactory level of attainment when selecting ideal set of goals:** Choosing ideal set of goals from multiple goals required negotiation and comparison of different priorities and identify best solution using varying number of resources and interests.

Stage 1 of the conceptual process is concluded here. In the next section, the design logic is presented which is based on the above requirements. A design decision is formed for each requirement in the design logic, followed by a discussion.

5.4 Design Logic

The requirements discussed in previous section are worked out as design logic, which serves as the second stage of the conceptual process. A concise outline of the design decisions is given in Figure 5.2. Following design logics are prepared fulfil each requirements.

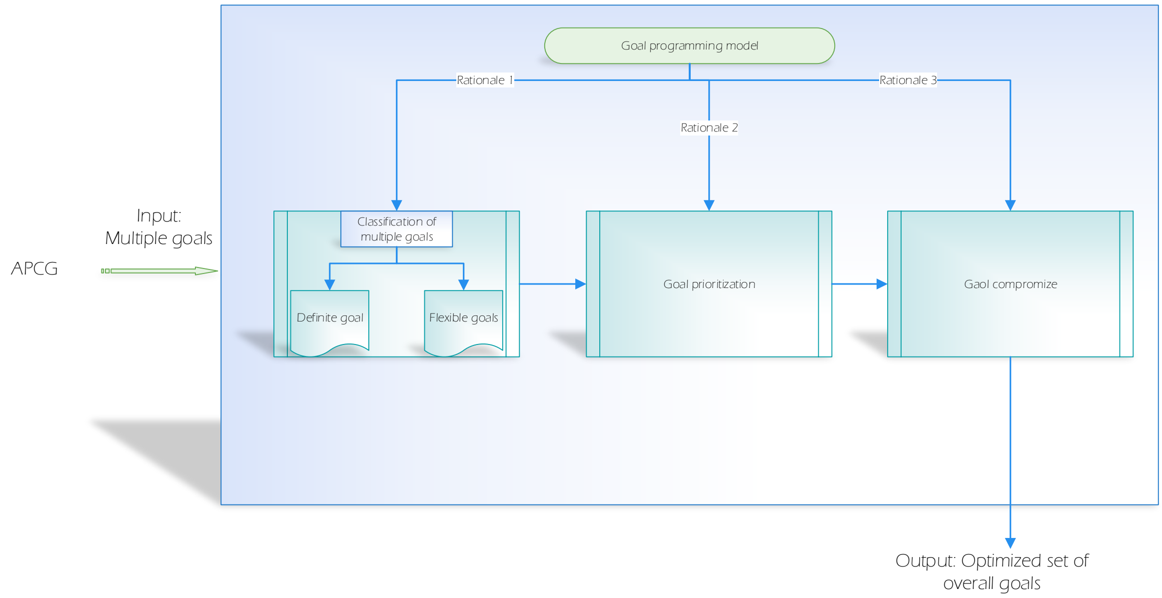

Figure 5. 2 Design decisions Outline

Design logic 1

The overall multiple goals are categorized into definite and flexible constraints, where definite constraints are the rigid restrictions on the goals such as carbon absorption practices should be implemented by APCGs. Flexible constraints are those conditions that are desirable but not must, such as the stability of APCGs or cost reduction objectives. This is the logic used to satisfy the requirement for multiple goals classification within the APCG (Req. 1).

Design logic 2

Some goals are more critical than others; hence, it is important to prioritize them. An approach is proposed which utilizes a goal-ranking technique used by Multiple Criteria Goal Programming (MCGP). In this step, goals are prioritized based on their relative importance. Hence second requirement for goal prioritization is met using this logic (Req. 2).

Design logic 3

If there are any conflicting goals, a selection is made from the conflicting goals. The concepts of goal negotiation techniques i.e. MCGP are utilized, which aim to provide the optimised goal solution considering different priorities, resource variation and gaps to find the best solution (Req. 3). These design logics are used in the development of the solution.

5.5 Theoretical Fundamentals

This section discusses the theoretical basis for the framework. This step is the third stage of the conceptual process. An algorithm is proposed for design decision and an in-depth analysis of the design decision is presented.

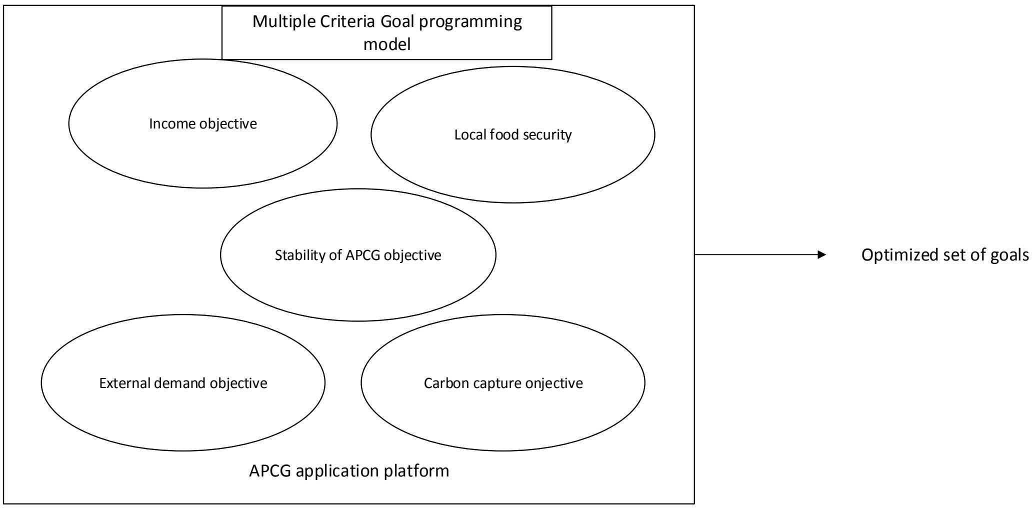

Figure 5. 3 Goal management- theoretical fundamentals

Multiple goals of the APCG are identified using the goal management framework. An objective programming model is developed using the MCGP concepts which is the key part of the system. The input for the goal management includes the diverse goals of the APCG, and the output is an optimised set of overall goals for the APCG. The optimised overall goals will help to ensure the ideal set of the goals.

Next, the development of the solution framework is described and discussed in detail.

5.5.1 Goal management

The goal management stage is responsible to attain ideal set of goals out of overall goals. The purpose involves solving diverse conflicting goals in the APCG to obtain best solution in terms of goals priority. The feature of MCGP [78] and an approach utilised in smart grid goal management[79] is used to design best possible solution for conflicting goals. Each and every identified objective is attached with a rank based on their priority. High rank objectives are treated as goals to work out first, and therefore attempts are made to find a solution which is close to the pre-ranking set of goals. Goal programming minimises the deviation between the theoretical goals and realistic achievements. These deviation can be both positive and negative, thus an objective function is used to minimise the deviations based on the relative importance of the goals.

Various areas has utilised goal programming model benefits such as environment, energy, smart grid, academic and health planning [14], and shows success in solving diverse conflicting goals. In this framework, we adapt MCGP techniques for our framework. Figure 5.4 presents the algorithm for the goal programming model, where the parameters and equations are explained in the following section. The model has six parts:

(i) APCGs goal recognition,
(ii) Summary of variables,
(iii) Objective classification,
(iv) Objective ranking,
(v) Goal equation formation, and
(vi) Generating objective functions.

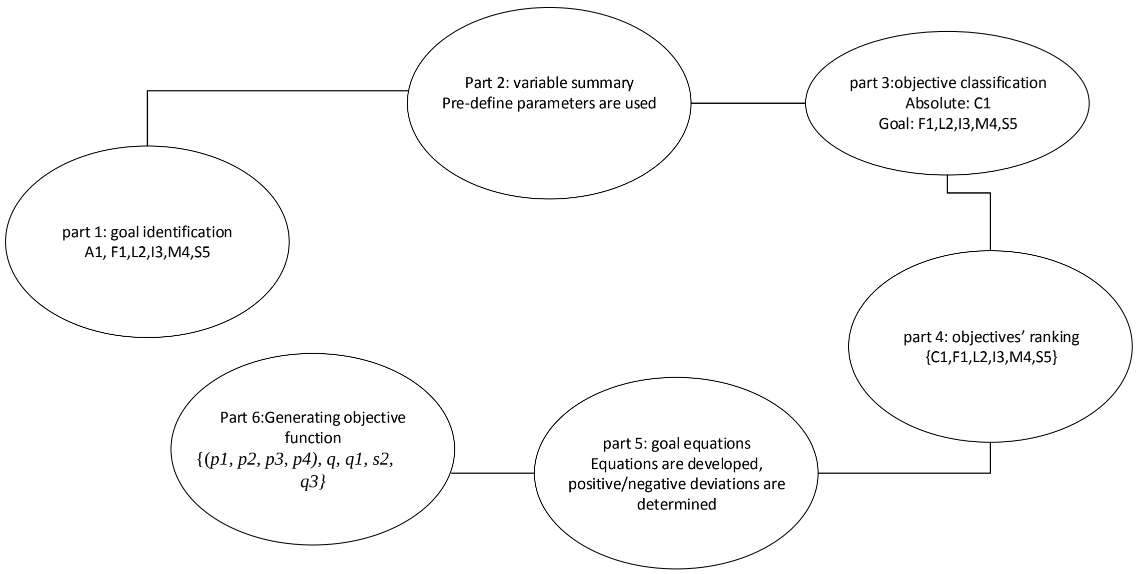

Figure 5.4 Goal method model

5.5.1.1 Part 1: APCGs goal recognition,

APCGs diverse goals are identified in this phase. These objectives are explained below.

1. ***Carbon content objective (C1)*:** The "carbon-capture objective" refers to the use of organic farming methods to maximize carbon capture, which will increase the carbon content which can be traded with external companies. More carbon capture will result in more carbon sequestration and less emission.
2. ***Food security within the network (F1).*** The goal is to secure the vegetable/fruit demand of local members within the APCGs. Realistically, some members within an

APCG may struggle producing sufficient quantity to meet their own consumption needs. Hence, food security of APCG members have been targeted.

3. ***Providing local food access to wider community (L2).*** With growing local food, APCGs can make locally grown vegetables available to the extended community such as external customers or supermarkets, greengrocers, and external consumers who are not registered with an APCG.
4. ***Income and Incentive objective (I3).*** The "income and incentive objective" focus is to earn income and incentive from selling surplus production of APCGs to vegetable/fruit buyers and trading carbon tokens with industries.
5. ***Maintenance cost reduction objective (M4).*** This goal refers to reducing the cost of APCGs maintenance over time. For example, "maintenance cost" may represent the one time cost to build APCG platform and maintaining the database and transaction records etc. Cost related to collection and distribution of products/vegetation from members, to stores etc. Additionally providing benefits to the members may require a payment gateway which may incur cost.
6. ***Stable APCG objective (S5).*** The increase in the number of active APCG members, that is, those who dynamically participate in the production-sharing or carbon-sharing network, is a "stability objective".

5.5.1.2 Part 2: Summary of variables,

In order to use MCGP all variables and their deviations are identified. For APCG the idea is to identify variables and summarize their deviations to achieve ideal set of goals. The production amount and carbon tokens generated by each group will be counted as variables and maximizing/minimizing the value is considered as deviation.

5.5.1.3 Part 3: objective classification

The objectives are classified as definite and flexible constraints based on the previous objectives (part 1). At this point, the "definite goals" are outlined as mandatory requirement on the variables, whereas the "flexible goals" are outlined as the objectives nice to have but not necessary [80] . The classification is of goals are as follows:

I. **Definite goals:** Maximum carbon capture objective (C). For example, the APCG's base is environmental sustainability. Thus, ecological methods must be used for APCG production.

II. **Flexible goals:** Goals such as local food security (F1), extended community and customer demand objective (L2), income & incentive objective (I3), maintenance cost objective (M4), and stability of APCG (S5). Refinement of these goals helps in achieving the ideal goal set, which would benefit APCG. The variables summaries is defined as: maximum C1, minimum F1, minimum L2, minimum I3, maximum M4, and minimum S5; these are termed "expected values" in the goal programming model.

5.5.1.4 Part 4: Objective ranking

To make sure important goals met first, the priorities of the goals have been assigned. This step discusses ranking out the goals by assigning a weight (or rank) to each goal. As mentioned earlier, goals can be mutually exclusive; i.e. one goal may be achievable at the expense of another. This makes it critically important to assign weights to the goals, so that least important goals are only met after the important ones.

Keeping local network food security (F1) as priority, total goal set can be determined as 4!. Thus in total 24 structures will be formed such as F1L2I3M4S5, F1L2M4S5I3… F1S5M4I3L2.

5.5.1.5 Part 5: Goal equation formation

Mathematical relations are developed in this section for the definite and flexible goals. Equations are as follows-

(i) Carbon capture Objective (C): Organic farming methods should be used for APCG produce to increase the carbon token value.

***(ii) Food security local demand objective (GC1)*:** Satisfying food security of APCG should be focused. Thus, the purpose of this goal is to minimise the negative deviation from the quantity of surplus production of each APCG.

Let A_{pi} E_i be the extra production produced by i^{th} APCG, k_0 and l_0 be negative and positive variance respectively, and t be the number of APCGs; then the equation for food security local demand objective (F1) would be:

$$A_{pi} \times E_i \geq 0; \forall i \leq t$$

$$A_{pi} \times E_i + k_0 - l_0 = 0; \forall i \leq t \qquad \text{Equation 5.1}$$

Considering 4 APCG groups for this framework, 4 equations will be formed (*m=4*) for each group;

$N_{p1} \times E_1 + k_1 - l_1 = 0; \quad \ldots\ N_{p4} \times E_4 + k_4 - l_4 = 0;$

(iii) Local community demand objective (L2): The purpose of L2 is to minimise the negative variance of the total surplus production of all APCG. Assuming requirement from external supermarket is R. And positive and negative variance be *s* and *q*, respectively; then the equation will be formed as

$\sum_{i=1}^{m} E_i \times A_{pi} \geq R$

$\sum_{i=1}^{m} E_i \times A_{pi} + q - s = R$ Equation 5.2

(iv) Income & Incentive objective (I3): Obtaining higher income is another requirement of the framework. The minimum income expectation of the i^{th} APCG be I_i, and positive and negative variance be *q1* and *s1* respectively; then the equation for this objective will be minimizing negative variance:

$\sum_{i=1}^{n} I_i \times E_i \times A_{pi} \geq I$

$\sum_{i=1}^{n} I_i \times E_i \times A_{pi} + q1 - s1 = I$ Equation 5.3

(v) Maintenance cost objective (M4): Let say the maintenance cost allowances be M, and the positive and negative variance be q2 and s2, respectively; equation for the maintenance cost objective (GC4) is obtained with Equation 5.6, where C_i is the coefficient, represents the cost rate of i^{th} APCG.

$\sum_{i=1}^{n} C_i \times E_i \times A_{pi} \leq M$

$\sum_{i=1}^{n} C_i \times E_i \times A_{pi} + q2 - s2 = M$ Equation 5.4

(vi) Sustainability objective (GC5): Let P be the minimum number of prosumers who are participating in APCG, and positive and negative variance be q3 and s3, respectively; then, the formula for the sustainability objective (G5) would be:

$\sum_{i=1}^{n} A_{pi} \geq P$

$\sum_{i=1}^{n} A_{pi} + q_3 - s_3 = P$ Equation 5.5

5.5.1.6 Part 6 Development of objective functions

Finally the objective function of each goal is formulated and, best possible solution is formed by minimizing the deviations from each goal. The objective functions here are the [(k1,k2,k3,k4*), q, q_1, s_2, q_3]*. *Partitioning algorithm* is used to solve this linear goal programming problem.

5.5.1.7 Goal programming solution

As discussed previously, 24 priority goal structure sets are identified along with different ranking order. The partitioning algorithm is utilized as a solution here, in order to solve the linear goal programming problem [79]. The solution working principle implies on the definition of priority structures which implies that higher-order goals must be optimised before lower-order goals are even considered. The solution procedure is shown in Figure 5.5 which consists of solving a series of linear programming sub-problems by using the solution of the higher-priority problem solved prior to the lower-priority problem. All the sub-problems assigned to a higher priority goals are solved first using the partitioning algorithm. The ideal tableau for this sub-problem is then examined for alternative ideal solutions. If none exists, then the present solution is ideal for the original problem with respect to all the priorities.

The algorithm then substitutes the values of the parameters for the flexible goals of the lower priorities to calculate their satisfaction levels, and the problem is solved. However, if alternative ideal solutions do exist, the next set of flexible goal and their objective function terms are added to the problem. This brings the algorithm to the next sub-problem in the series, and the optimisation resumes. The algorithm continues in this manner until no alternative ideal exists for one of the sub-problems or until all priorities have been included in the optimisation [78, 80].

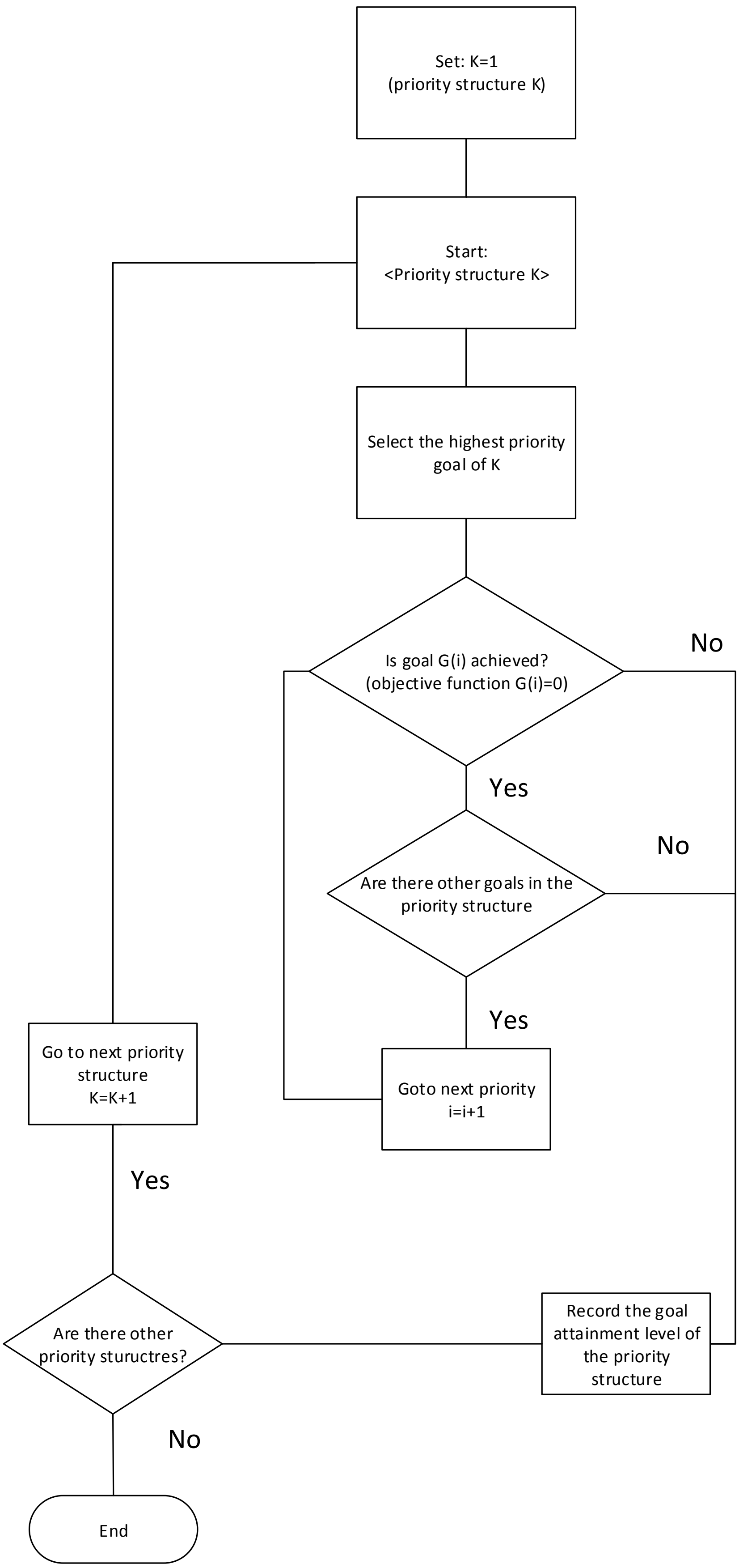

Figure 5.5 Partitioning algorithm for APCG's goal management

Goal management problem provides the best solution by comparing the achievable set of goals when compared to the predetermined goals. Additionally the identification of the necessary alterations to parameters are explained well in order to achieve all the goals in different priority structures.

5.6 Prototype verification of the goal management framework

Verification of the framework is presented in this section. The multiple goal management is modelled in an APCG network as a linear goal programming problem. The solution is developed using LINGO, and is discussed in the following sub-section.

First, the experimental settings are discussed which is used for solving the goal programming model.

5.6.1 Experimental setting

Table 5.1 shows some of the parameters for the goal programming problem that are obtained based on the generated prosumers data; lemon production parameters are calculated based on the Australian average production quantity and variety (http://seasonalfoodguide.com/australia-general-seasonal-fresh-produce-guide-fruits-vegetables-in-season-availability.html). The seasons are also picked up using Victorian weather conditions and as real data could not be accessed or found. Victorian suburb population is considered as sample population data. Here, we take the four APCG groups discussed previously in APCGs concept framework (Chapter 3).

Table 5.1 Parameters for goal programming model

Prototype Parameter Yearly	Value
Average production	
Agro-prosumer community group 1 (APCG 1)	129
APCG 2	255
APCG 3	385
APCG 4	690
Available number of prosumers	
APCG 1	59
APCG 2	20
APCG 3	20
APCG 4	37

*Suburb demand (calculated for 1 suburb)	45,000
Income rate (assumed weights) APCG 1:APCG 2:APCG 3:APCG 4	1:3:6:9
Total expected Carbon token count	10
Total expected income (assumed) **	$ 11,650
Cost rate (assumed weights) APCG 1:APCG 2:APCG 3:APCG 4	1:2:3:4
Total budgeted cost constraint (assumed) ***	$ 1,000
The percentage of overall participations sustainability (Ns)	90%

To ease the calculations, local food security demand objective is chosen top priority and keep it the same for all the possible solution structures. Thus reducing total possible solutions to 4! i.e. 24 structures. The different priority structures are formed, where the position of the characters (“F1”, “L2,” “I3,” “M4” and “S5”] shows the priority order of the different goals. LINGO-32 is used to program the algorithm. The observations and results obtained by solving the goal problem in LINGO is presented in next section.

5.6.2 Observation, results and discussion

The solution is illustrated in this section. The goal management solution picks the best solution by predicting the division of the objective function according to the process priority level and the sequential solution of the resulting mixed integer linear programming model. The solution obtained at each priority level is used as a constraint at the lower level. The general examples discussed here are intended to illustrate the model's applicability to the problem of practical dimensions.

For instance, I3 on priority sets the objective function for I3 to 0, but increases objective function for L2 to 35564.50. When L2 is set on priority M4 successfully met but I3 increases to 11650. When setting L2 on priority increases the I3 to 11651 and M4 to 84446. Setting M4 achieve just for M4 but does not met for L2 and I3. Same applies for S5. So, putting I3 on top achieves the most except for S5. Hence, making S5 the next priority will help to achieve all desired goals. Putting L2, I3 and M4 objective function together on same priority help achieve the best. Therefore, the negotiated priority set of goals are C1F1L2I3M4S5 which is illustrated in Table 5.2.

Table 5.2 Negotiated set of ideal goals

Goals	Details	Value
F1	local demand of APCG	5,440
L2	Maintain Suburb demand	45,000
I3	Maximise the total expected income	$ 11,650
C1	Maximise carbon token	8-10 token/year
M4	Minimise the cost	$ 1,000
S5	Sustainability	90%

5.7 Conclusion

In this chapter, a framework is presented for agro-prosumer community group's goal management. MCGP techniques are utilised to develop the goal management framework. The proposed solution determines the diverse, conflicting goals within the network, and prioritizes them based on their influence. It also negotiates the goals to obtain the ideal set of goals for the network. Thus the chapter fulfil the third component of the APCG framework and motivates agro-prosumers to work on their interests and achieve network goals, while earning incentives and managing the best possible order.

Chapter 6

6 Carbon Assessment and Blockchain Framework for Carbon Sharing

This chapter provides:

- An introduction to prosumers' carbon assessment and carbon token framework
- A framework for APCG implementation using blockchain, discussed in detail.
- Verification of the proposed framework for APCG carbon token generation.

6.1 Introduction

In the last chapter, we proposed a definition and allocation framework to manage the overall and mutual goals of community groups.

One of the key motivating goal for APCGs is income where the prosumers earn incentives when selling their surplus produce or trading carbon capture with industries and organizations. Carbon capture can be shared between the prosumers' group and industries in a digitalised form, that is, a carbon token. It is important to safeguard the quality of products by ensuring traceability in the framework. Similarly, for carbon token, it is desirable to keep a track of fair transactions, and to avoid double spending (selling tree-related carbon multiple times to earn carbon credits). Thus, a transparent and accountable framework is required to implement APCG.

Blockchain is an innovative decentralised technology whereby information is stored in interconnected blocks. In each block, transactions are stored, and information about transactions is secured through cryptography. Blocks can be accessed publicly, and stored information is immutable. Hence, blockchain is appropriate for the implementation of an APCG framework as it allows transparency and traceability.

This chapter presents an APCG carbon assessment framework and a prototype APCG using blockchain. The main objective of this chapter is to discuss carbon evaluation and the design phases of an APCG framework using blockchain.

In Section 6.2, we provide an Outline of carbon assessment and carbon token framework. In Section 6.3, the requirements for this framework are discussed, followed by the design logic in Section 6.4. In Section 6.5, we discuss the theoretical fundamentals of the framework, which is verified in Section 6.6. Section 6.7 describes the prototype implementation using blockchain technology. Section 6.8 concludes the chapter.

6.2 Outline

With global food systems, a question to food insecurity and food safety is growing. According to the Food and Agriculture Organisation, food insecurity is defined as lack of secure access to sufficient and nutritious food, which is essential for normal growth, development and a healthy lifestyle. Food insecurity has several factors such as increasing food prices, low income, lifestyle, limited access to services etc. Food insecurity is high in developing countries but it is also a problem in developed countries (https://sei.sydney.edu.au/opinion/can-urban-agriculture-reduce-food-insecurity-urban-poor/).

An agro-prosumer community group offers a platform that provides dual benefits. Firstly, it provides a food security system to local community and market by forming a vegetable/fruit sharing network. Secondly, at the same time, group members are contributing to environmental sustainability while storing carbon naturally through ecological farming practices. Carbon sequestration is a biological process of sequestrating carbon di oxide in soil (Figure 1). As a solution to reduce increased CO_2 and GHG emissions, the APCG framework can encourage this biological process by engaging agro-prosumers in carbon token generation which can be traded with companies to offset their carbon production. Agro-prosumers can generate carbon tokens depending upon the consumption of total amount of carbon content during the crop growing process. In this section, a solution for carbon assessment and carbon token framework is proposed.

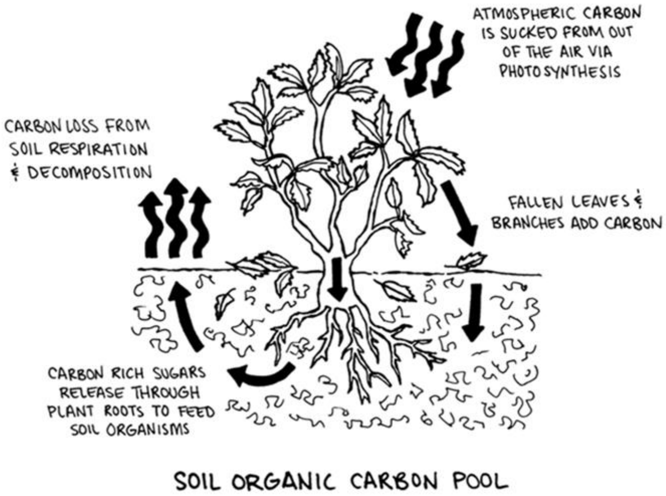

Figure 6.1 Carbon sequestration biological process

The solution is presented in two parts. In the first part, the assessment of agro-prosumers' carbon is done using carbon capture and carbon emissions parameters. Then, the output of the carbon assessment is used to generate carbon tokens. Carbon calculations for all plants and trees is a cumbersome task, and this is the reason this study has the limitation of using trees for the assessment of agro-prosumers' carbon. In order to perform carbon assessment accurately, a carbon calculation method available on Internet (https://www.ecomatcher.com/how-to-calculate-co2-sequestration/) is used and DEFRA's report (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69282/pb13309-ghg-guidance-0909011.pdf) is also used to evaluate carbon emissions for each agro-prosumer. For each agro-prosumer, carbon absorption is computed and their carbon emissions are evaluated. Finally, the carbon captured is transformed into the carbon token. The carbon token can be shared/traded within the network or with external organizations by using blockchain technology as a platform for carbon token transactions or to trace the food produces. Figure 6.2 presents an Outline of the framework.

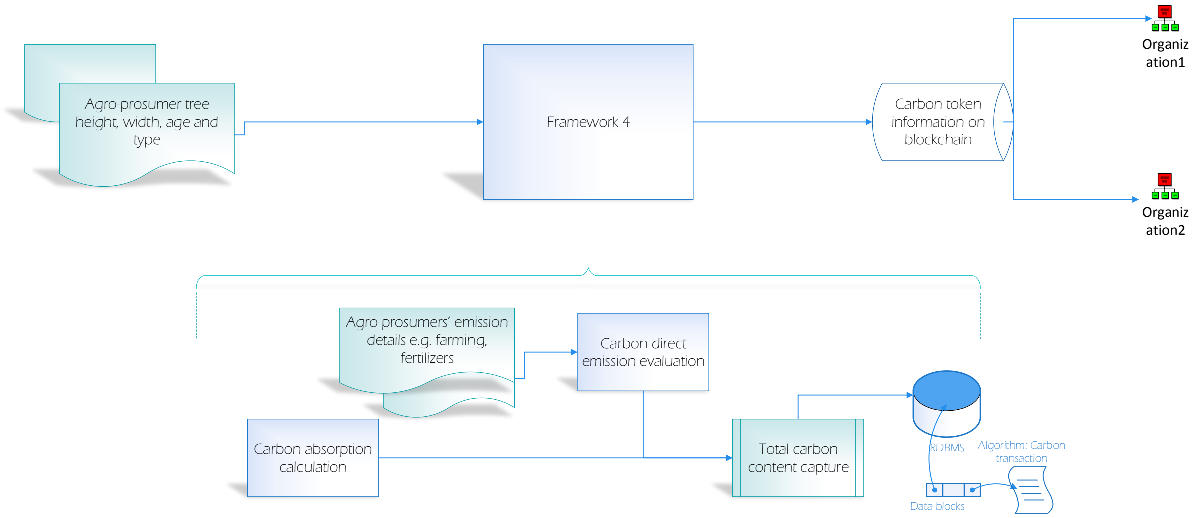

Figure 6.2. Carbon assessment framework and APCG design

In the next section, the proposed solution is discussed in detail.

6.3 Requirements

The requirements for the proposed solution are discussed in this section. In this first stage of the conceptual process, the requirements are stated and prioritized. The following are the requirements of the solution, where a three-step method suggested by DEFRA (UK report) is utilized for carbon capture.

6.3.1 Identifying carbon absorption

The total amount of carbon captured by APCGs is estimated in order to determine carbon absorption and convert this to carbon tokens that can be traded with industries. Trees are a long-term solution for the carbon sequestration process (because of their longevity and deep roots); thus, carbon absorption by trees is considered in this study. Therefore, information such as number of trees, their respective age, height and diameter, and type/variety can help to determine the amount of carbon a prosumer can sink. Thus, tree details are taken as input to identify carbon absorption for each prosumer; all outputs for all prosumers in an APCG are combined to arrive at the number of tokens that, collectively, an APCG has acquired.

6.3.2 Calculating carbon emissions for prosumer groups

With growing need for carbon emission reduction, APCGs' own sustainability requirements are prioritised. Prosumers contribute to carbon emissions in two ways: direct emissions result from the use of tools/machine, and fuel combustion etc.; indirect emissions are produced by chemicals, inorganic fertilisers etc. The direct carbon emission for each member is evaluated in this study and combined to total the group's total emissions.

6.3.3 Evaluating carbon capture and deducting prosumers' carbon footprint

APCGs' total direct emissions are deduced from the total carbon absorption to measure total carbon content captured. Thus, it makes APCGs sustainable and surplus carbon tokens can be traded with external companies and organisations to offset their carbon emission. The total carbon content capture is transformed electronically into tokens for trading purposes.

The stage 1 of the conceptual process is concluded here. In the next section, the design logic is presented, which is based on the aforementioned requirements. A design decision for each requirement is made in the design logic; also discussed is the way that each requirement is met.

6.4 Design logic

The second stage of the conceptual process is the design logic. This section provides a design logic that meets all the requirements listed above. The following design decisions are proposed in order to address each requirement. A concise Outline of the design decisions is shown in Figure 6.3.

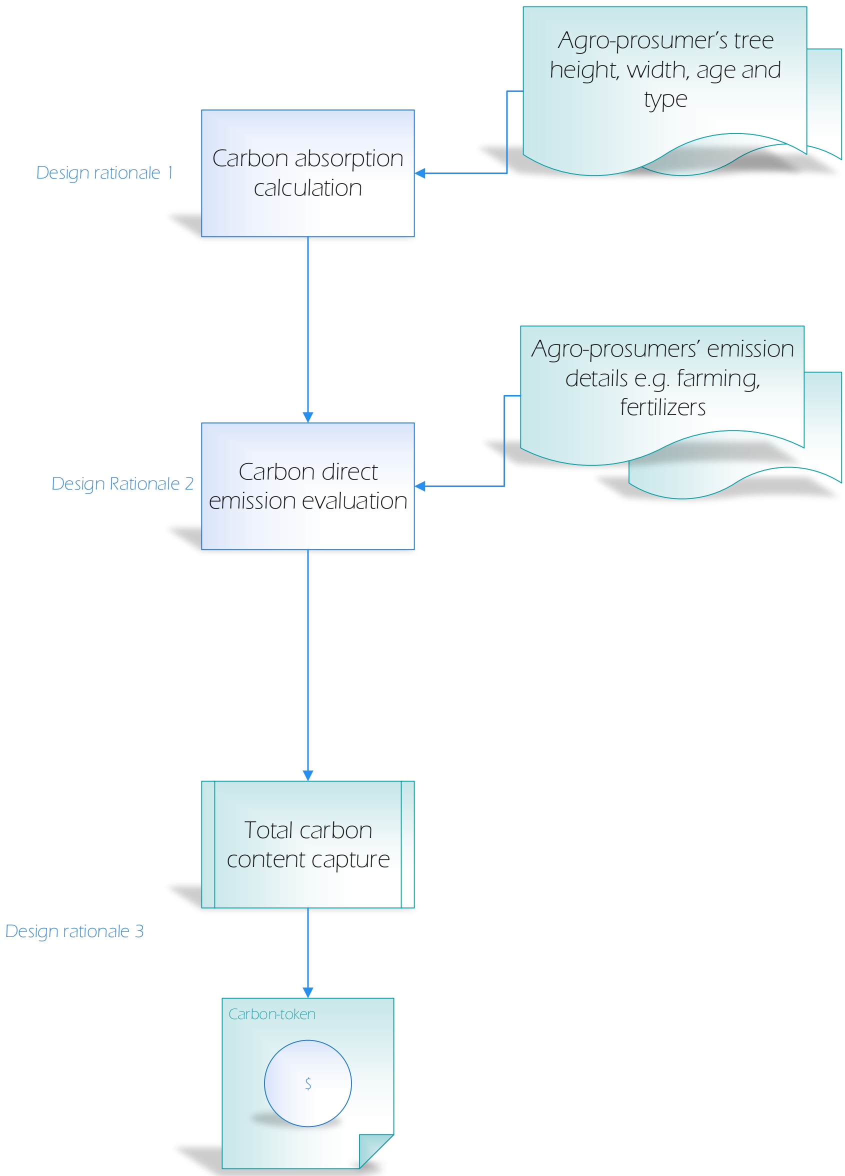

Figure 6.3 Design Outline of carbon framework

Design logic 1

Identification of captured carbon for each agro-prosumer a tree's carbon is calculated using a carbon calculator where inputs such as type of vegetation, tree's height from ground and diameter, are taken to calculate the carbon capture. Hence, for each APCG, the total number of trees and their respective carbon content computation can be combined to obtain the total amount of carbon captured by each group and by the whole APCG.

Design Logic 2

Identification of carbon emission activities for each agro-prosumer

Farming practices affect the amount of carbon absorption. For instance, organic and traditional methods produce better carbon content than do modern farming activities. Hence, to make APCG sustainable, the amount of prosumers' carbon emission must be calculated first. The farming activities that involve the use of tools/machines and chemical fertilizers must be determined for each member within a group. The amount of carbon emission from the farming practices of each prosumers is calculated and combined with those of other prosumers to arrive at the total emissions for the group.

Design Logic 3

Evaluation of carbon token presents. The total amount of carbon captured is computed by deducing the total direct emissions of APCGs from the total amount of carbon captured by APCGs. Thus, the total amount of carbon captured can be converted into carbon tokens and allocated for trading purposes.

In the development of the solution for a carbon token system, the above design decisions should be considered. This concludes stage 2 of the conceptual process. The next stage establishes the theoretical fundamentals for the proposed solution.

6.5 Theoretical fundamentals

This section discusses the theoretical fundamentals for the carbon token framework. This is the third stage of the conceptual process. A carbon calculator method is used to calculate the total amount of carbon captured. Captured carbon is ultimately converted into digital carbon tokens. This section is divided into two parts: (i) Part 1: carbon estimation, and (ii) Part 2: Carbon token framework.

For carbon estimation, carbon absorption parameters are used to estimate the total amount of carbon absorbed by each agro-prosumer community group. A carbon calculator is developed as an automated program to calculate captured carbon. The input parameters for the carbon calculator includes carbon absorption and emission details for each member of the group. This results in the total amount of carbon captured by APCGs to trade with industries.

In part 2, a carbon token framework is developed to transform the total amount of carbon into digitalised form. A carbon token, which is a digital form of total carbon dioxide absorbed by an APCG, can be traded with industries and organizations.

6.5.1 Part 1: Carbon calculations

The carbon calculation phase involves the computation of the total amount of carbon captured by agro-prosumer community groups. A DEFRA method is used to estimate the total amount of carbon captured by APCGs. This is converted into digital carbon tokens. It is also important to make APCGs self-sustainable and responsible for the carbon footprints of their own group.

A carbon calculator is used to compute the amount of carbon absorbed by a tree. Later, the DEFRA method is used to measure, evaluate and convert the carbon dioxide equivalent and calculate the total emissions. Finally, emissions are deducted from absorption to estimate the amount of captured surplus carbon. This part has three steps: (i) carbon capture estimation (ii) carbon emission evaluation and (iii) carbon token formation.

6.5.1.1 Carbon capture estimations

This step estimates the carbon sequestration of trees for APCGs using the carbon calculator. For APCGs, carbon sequestration is computed for lemon fruit trees. Fruit trees are said to sequestrate more CO_2 (http://www.co2science.org/subject/t/summaries/treesfruit.php), because of the fact that as the air's CO_2 content increases, fruit trees have greater photosynthesis and higher biomass production, regardless of soil moisture conditions. Consequently, greater amounts of carbon are likely to be sequestered in the woody trunks and branches of these species. In addition, fruit yields may increase as well, thus contributing more to APCGs' production.

However, the rate of carbon sequestration depends on the growth characteristics of the tree species, the conditions for growth where the tree is planted, and the density of the tree's wood. It is greatest in the younger stages of tree growth, between 20 to 50 years. To roughly estimate the amount of CO_2 sequestrated in a given tree, a four-point method is used.

1. Determine the total (green) weight of the tree

The green weight is the weight of the tree when it is alive. Firstly, the green weight of the above-ground weight is calculated as follows:

$W_{above\text{-}ground} = 0.25\, D^2\, H$ (for trees with D<11)

$W_{above\text{-}ground} = 0.15\, D^2\, H$ (for trees with D>11)

Where, $W_{above\text{-}ground}$ is above-ground weight in pounds, D is diameter of the trunk in inches and H is the height of the tree in feet.

The root system weight is about 20% of the above-ground weight. Therefore, to determine the total green weight of the tree, multiply the above-ground weight by 1.2:

$W_{total\ green\ weight} = 1.2 * W_{above\text{-}ground}$

2. Determine the dry weight of the tree

The average tree is 72.5% dry matter and 27.5% moisture. Therefore, to determine the dry weight of the tree, multiply the total green weight of the tree by 72.5%.

$W_{dry\ weight} = 0.725 * W_{total\ green\ weight}$

3. Determine the weight of carbon in the tree

The average carbon content is generally 50% of the tree's dry weight total volume. Therefore, to determine the weight of carbon in the tree, multiply the dry weight of the tree by 50%.

$W_{carbon} = 0.5 * W_{dry\ weight}$

4. Determine the weight of carbon dioxide sequestered in the tree

CO2 has one molecule of Carbon and 2 molecules of Oxygen. The atomic weight of Carbon is 12 (u) and the atomic weight of Oxygen is 16 (u). The weight of CO2 in trees is determined by the ratio of CO2 to C is 44/12 = 3.67. Therefore, to determine the weight of carbon dioxide sequestered in the tree, multiply the weight of carbon in the tree by 3.671.

$W_{carbon\text{-}dioxide} = 3.67 * W_{carbon}$

6.5.1.2 Carbon emission evaluation:

According to a DEFRA report (www.defra.gov.uk), carbon emission is estimated using a three-step process: calculating different types of emissions, identifying the quantity used and then combining them to compute total emissions. The carbon emission evaluation adapts DERFA concepts to evaluate total carbon emission for prosumers who are engaged in community farming.

- Step 1: Calculating direct emissions: Direct emissions are those released due to operating own vehicles, heating or boilers. In the case of APCGs, direct emission is related to tilling, machine use etc.
- Step 2: Calculating indirect emissions: Indirect emissions result from individual actions, but where the sources of these energies are not owned by them. Examples are materials, fertilisers, goods etc. For APCGs, indirect emissions come from the use of chemical or inorganic fertilizers, seeds, soil etc.

Step 3: Convert to carbon dioxide equivalent and calculate total emissions: The DEFRA report (pg. 17) notes that various activities or fuels have different Emission Factors, reflecting their level of pollution. There are six different GHGs: carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and sulphur hexafluoride (SF6). Different activities and fuel can release one or more of them, each with its EF. The following formula given in the DEFRA report (pg. 18) is used to calculate emissions

Activity data (d_1) x Emission factor (ef_1) = GHG emission (E_1)

Activity data (d_2) x Emission factor (ef_2) = GHG emission (E_2)

Activity data (d_n) x Emission factor (ef_n) = GHG emission (E_n)

Total Emission E or CO_2e;

$CO_2e = E_1+E_2+\ldots+E_n$

Hence, to calculate emissions from the use of a resource or service, multiply its total use in a year by the emission factors of all pollutants produced by it. Finally, all emissions are combined as total emissions of carbon dioxide equivalent.

6.5.1.3 Carbon token formation

The total amount of carbon dioxide equivalent (CO_2e) is deduced from total carbon absorbed ($W_{carbon\text{-}dioxide}$) to calculate total carbon captured.

Thus, the total amount of carbon captured would be:

Carbon-Captured= $W_{carbon-dioxide}$-CO_2e

Now, 1 tonne of CO_2 is equivalent to 1 carbon credit.

The captured carbon- is converted into tonne unit and carbon credits are formed using one carbon credit which is equivalent to 1 tonne of CO_2. These carbon credits will be converted into digital carbon tokens using a Python code that utilizes Numpy and Pandas library functions.

6.6 Carbon assessment and token prototype

6.6.1 Experimental setting

For verification and validation of the carbon token framework, the solution framework is simulated. Firstly, the experimental settings for the proposed framework re discussed.

The setting here is a basic set-up for the proposed framework examination. To verify the proposed algorithm, one APCG group is considered for evaluation of their carbon content based on the lemon trees. For dataset generation, production behaviour along with consumption pattern is used from chapter 3 for APCG1. Microsoft Excel is used to run the algorithm and do the computation.

The prototype parameters are shown in Table 6.1 below.

Table 6.1 Prototype Parameters

Parameters	**Value**
Lemon variety	Eureka, Lisbon and Meyer
Number of trees per prosumer	1
Tree age (assumed)	3-6 years
Height	4-6m for Meyer and Eureka, 6-8m for Lisbon
Diameter	4-6m for Meyer and Eureka, 6-8m for Lisbon
Farming method	Organic

To simulate the framework, APCG1 with 15 prosumers is considered here. The total number of trees per prosumer is restricted to 1 for easy data generation. As APCGs provide a platform for sharing both carbon and produce, it is assumed that the overall goal of APCG1 is to increase carbon capture by practising organic farming methods. Thus, an organic farming method is taken as a parameter for the prototype.

6.6.2 Results and discussion

In this section, the proposed framework's theoretical aspects are evaluated. As discussed in the theoretical fundamentals section (6.5) the framework has two phases: carbon calculations and carbon token. The carbon calculation of all prosumers' carbon capture is done using DEFRA's algorithm.

For APCG1, as a key goal is to use ecological farming methods, direct and indirect emissions show 0 contribution to CO2e. Hence, the total carbon captured is equivalent to the total carbon absorbed by the lemon trees. Table 6.2 below shows the calculation done for all prosumers in APCG1. The height and the diameter of the tree are converted into feet and inches respectively. DEFRA's equation is used to calculate the carbon absorption in pounds.

Table 6.2 Carbon captured by APCG 1 prosumer group

Variety of Lemon	Height and Di	Height in feet	Diameter in inches	Deraf's calculation
Lemon-Lisbon	6	19.68	11.8	656.1991885
Lemon-Eureka	4	13.12	9	254.4869016
Lemon-Lisbon	7	22.96	20	2199.26952
Lemon-Eureka	5	16.4	5	98.181675
Lemon-Eureka	6	19.68	15	1060.36209
Lemon-Meyer	4	13.12	6	113.1052896
Lemon-Lisbon	8	26.24	8	402.1521408
Lemon-Meyer	4	13.12	10	314.18136
Lemon-Lisbon	5	16.4	3	35.345403
Lemon-Eureka	6	19.68	7	230.9232996
Lemon-Eureka	5	16.4	9	318.108627
Lemon-Eureka	5	16.4	11	475.199307
Lemon-Meyer	4	13.12	12	452.4211584
Total carbon-absorption in pound				**6609.93596**

Total carbon absorption is converted into metric tonne units to calculate carbon token. ~3 tonne of carbon is stored by one APCG group for 1 year. Similarly, the carbon token can be computed for each APCG group and combined. Thus, the carbon tokens can be collectively shared by APCG to organizations or government bodies to offset their carbon.

To implement the carbon token framework, blockchain technology is used. The algorithm can be implemented and utilized either for carbon token or produce transactions as an asset. The implementation design is shown in the next section.

6.7 APCG implementation using Blockchain

APCG is a framework that aims to offer a community network sharing group. APCGs not only satisfy local food demand but also encourage environmental sustainability by producing carbon tokens. Generally, a network sharing group is operated by a trusted centralized authority: intermediaries that require a system to maintain transactions records and safety with high integrity, reliability, and security. However, a central system often cannot guarantee the security, safety and quality of agricultural products. Therefore, it is desirable to build a robust platform for APCGs that provides security, safety and transparency in the network. Thus, a technology which utilises distributed ledger technology (DLT) to keep information and records unchangeable seems a promising solution for APCGs [13, 18, 26].

The blockchain is basically an immutable and decentralized, shared public ledger of transactions which allows participants to keep track of transactions without central record-keeping tasks. Blockchain utilizes DLT, and ensures that the stored information is impossible to hack, erase or alter. Hence, this technology makes the information transparent and safe compared to conventional systems for information storage (www.hospitalitynet.org). The proposed solution focuses on the utilization of smart contracts to govern and control all interactions and transactions among all the participants involved in the APCG network.

This chapter focuses on the utilization of blockchain technology to create smart contracts to control all transactions and information exchange within the APCG network and with the organization/external buyers and APCG. All transactions are recorded and stored in the blockchain's immutable ledger, thus providing a high level of transparency and traceability for the APCG. However, not all the information is saved in the ledger, the framework takes advantage of RDBMS and blockchain ledger storage to build an economical model. Next section discusses the design steps to implement APCG using blockchain concepts, however implementation of APCG using Hyper ledger or Ethereum is not in scope of this thesis.

6.7.1 APCG framework design:

Four frameworks for the establishment of an APCG were discussed in the previous chapter. There are two main nodes/entities in the network: (i) prosumer (ii) consumer/organization.

Each entity has a role, association and transactions with smart contract. Figure 6.4 shows the prototype Outline of the smart contract idea for APCG.

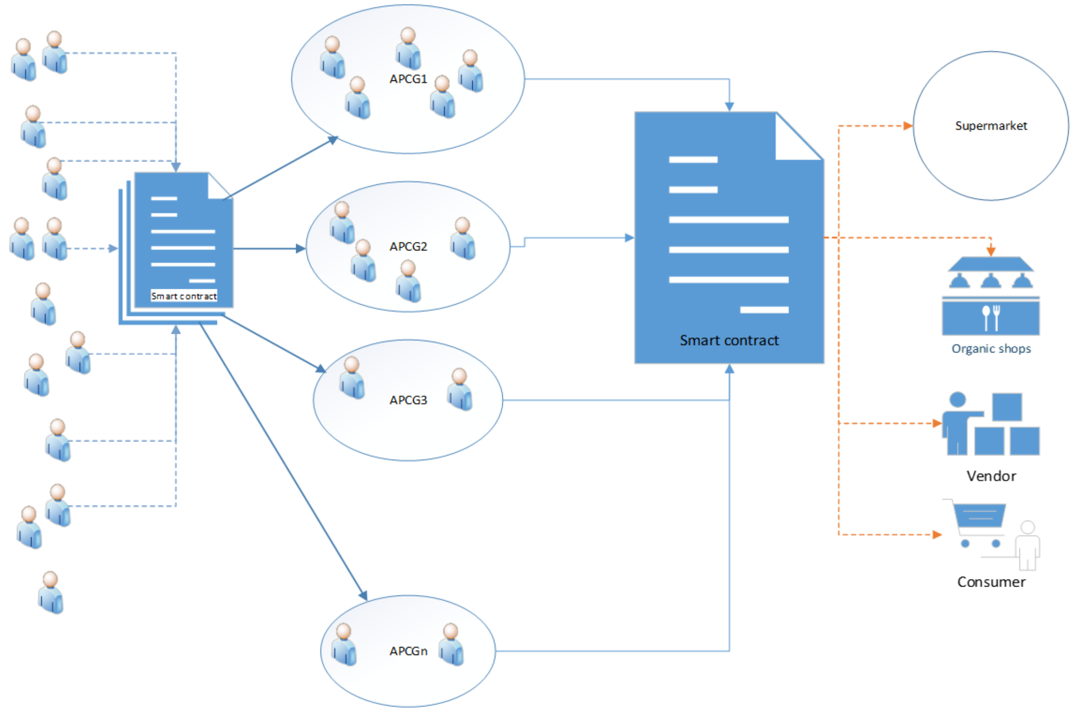

Figure 6.4 Outline of APCG prototype using blockchain

Figure 6.5 is a detailed entity-relationship diagram which shows the attributes, functions and relationship between the participating entities and the smart contract. The overall APCG framework comprises entity, data fields, data storage and transactions as explained below.

1) APCG definition and characterisation: Any individual can register as a prospective prosumer by providing basic personal and family details, which are stored as an Agro-prosumer dataset which will be part of the RDBMS data warehouse. There will be other datasets which will provide the property information and detailed crop information. There is no limitation on the number of properties a prosumer can propose for an AGRO business. The APCG evaluation matrix is performed at property level not at prosumer level. The various related attributes are captured in the Entity Relationship Diagram.
2) APCG prosumer Engagement: In order to analyse new prosumer production behaviour, the evaluation period production information is stored and the stability index is evaluated. This information is stored in APCG dataset which again is part of the relational database.

3) APCG goal management: The list of agro-prosumer group production characteristics is stored in APCG database, which is then used to define and negotiate among overall APCGs goals.

4) APCG carbon assessment and token: prosumer and consumer contracts along with property level carbon assessment will form a smart contract which will be stored on blockchain. Two smart contracts are used in this framework. One is between the prosumers and other is between the APCG network and external parties such as organisations/ buyers/retail stores. In blockchain, all data and records are digitally signed and attributed to a certain actor. Because the prosumer is responsible for production, the prosumer is the undisputed owner of this information and is responsible for any inaccurate or fraudulent information. Blockchain via smart contracts can be programmed in an automated way to impose penalties on any prosumer who acts dishonestly.

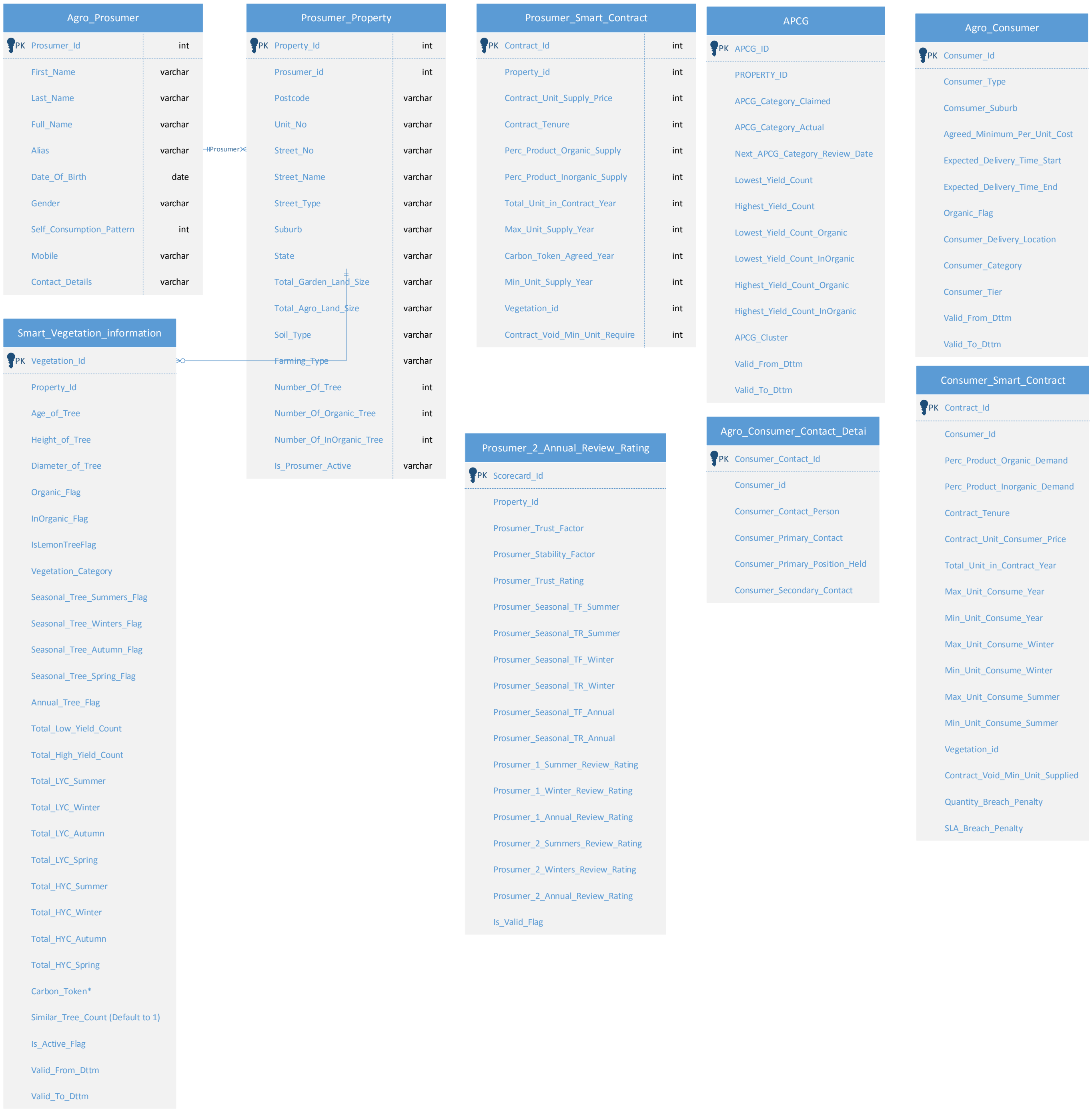

Figure 6.5 Entity-Relationship Diagram of APCG framework

6.7.2 Transactions

It is worth developing an algorithm that defines the working principles of our proposed blockchain-based approach. The following pseudocode is written for carbon token trade between APCG and an organization.

Algorithm 6.1 shows pseudocode for carbon-token transaction using blockchain concepts. Input for this smart contract will be total number of carbon tokens (carbon_token) generated by APCG, List of (property_id) of APCGs prosumers, Time period for (contract_tenure) carbon contract and price at which the (carbon_token_price) are sold.

Initially when transaction starts, contract_state is set as “created”, the status at APCG is set to “not sold”. Let’s say company XYZ is interested in buying APCG carbon_token at carbon_token_price. The status will be shown as “ready” for organization XYZ. The access for APCG carbon_token will be seized to XYZ. If XYZ is a registered company for APCG and if transaction amount from XYZ is paid to APCG then contract_state changes to “carbon_token_request_submitted”. APCG_network status changes to “agree to sell”. Finally a notification pops up as “Property_id sold to XYZ”.

Algorithm 6.1 Carbon-token transaction using blockchain

Carbon-token algorithm

Input: Carbon_token for APCG, property id, contract_tenure, carbon_token_price.
Contract_State: created
APCG network: not sold
Organisation XYZ: ready
Restrict access: XYZ
If XYZ= registered **and carbon_token_price=**paid **then**
Contract_State changes to: carbon_token_request_submitted
APCG_network changes to: agree to sell
Create a notification: Property_id_soldto_XYZ
End
Else
Revert transaction and show an error
End.

Similarly, a production traceability algorithm presented in Algorithm 6.2 is developed which uses production-id, seed and fertilizer brands and purchasing details along with agro-prosumers database to assure quality and safety issues.

Algorithm 6.2 Production traceability algorithm

Production tracebility algorithm

Input: prosumer_id, property id, contract_tenure, prosumer_id_price, seed brand, fertilizers brand
Contract_State: created
APCG network: not sold
Organisation XYZ: ready
Restrict access: XYZ
If XYZ= registered **and prosumer_id_price=**paid **then**
Contract_State changes to: prosumer_id_request_submitted
APCG_network changes to: agree to sell
Create a notification: Property_id_soldto_XYZ
End
Else
Revert transaction and show an error
End.

6.8 Conclusion

In this chapter, a prototype for a carbon token and assessment component of APCG framework is developed and a base for implementing the APCG framework is established. We explained and illustrated how agro-prosumer community groups' captured carbon information is computed and evaluated to obtain the surplus amount of captured carbon. This amount is then converted electronically for trade with companies and government bodies to offset their carbon production. This framework shows a novel and unique means whereby locals can participate in the carbon emission reduction program while engaged in local food production. In order to make APCGs transparent and traceable, blockchain technology is used to implement the model. The design phases are shown in detail and an algorithm is built to provide carbon token framework using blockchain.

This chapter contributes the last component of the APCG framework- carbon assessment. Phases to implement APCG framework using decentralized blockchain technology has been discussed to assist the proof of concept implementation.

As part of PhD work, APCG framework project work is concluded here. Next chapter 7 discuss a novel framework developed to evaluate users design perception using fuzzy regression techniques as part of another research work. For this research work, user design perception has been extracted from social media platforms and a case study has been presented to show the model capability. Chapter 8 summarizes doctoral internship research projects and work taken during the PhD period. A roadmap is developed to answer few current Machine Learning issues in health space, and provide a reference to be used while developing ML enabled tools/services. The road map is the main contribution from this chapter.

Chapter 7

7 A Social Media Analytical Framework Incorporating Fuzzy Regression for Affective Design

This chapter provides

1. The background of evaluating uncertainty present in users affective design data.
2. In-depth literature review of social media intelligence, big data, affective design and fuzzy regression
3. An outline of social media analytical framework methodology and fuzzy regression evaluation techniques.
4. Framework is demonstrated using a case study of automotive product and social media data.
5. Results, discussion and conclusion is presented at the end.

7.1 Introduction

In new product development, the factor which is attracting more interest other than the product's basic functionality is its affective or aesthetic quality [81]. Nowadays, users are more concerned about the colour, shape, look and feel of a product (i.e., the affective qualities of cars, smart phones etc.). Hence, affective design plays an essential role in new product development [82].

The consumers' voices play a critical role in the development or evaluation of a new product's affective design. Popular social media channels such as Twitter, Facebook and Instagram are significant sources of information and public content, and provide platforms for information exchange [42] . The social media data consists of users' opinion, comments, emotions and influences; thus, it contributes to a massive amount of knowledge sharing for various domains such as tourism, government, politics and product co-creation [43]. Social data contains useful information for identifying affective preferences of new products [44]. Using social media data, it is possible to extract information opinion regarding the affective satisfaction given by a product. Thus, the voices of customers can be extracted through affective design contents shared on social media, using social media data analytics [45].

Social media data is essentially a collection of consumer generated data, i.e., blogs, posts, comments, reviews and other forms of communication which are created by consumers. Since social data is influenced by personal feelings, tastes, and opinions [44], it is inherently ambiguous and uncertain. Given the highly competitive global market, it is essential to evaluate users' affective design perceptions along with the amount of uncertainty it contains. For instance, a new product is not just perceived as 'good' or 'bad'. People's judgement and opinion regarding affective design is expressed as 'I like it', 'I love it', 'I am not so sure', 'LOL' or even by using emoticons such as a thumbs up, smile, grumpy etc. Taking this into account, a significant need arises to estimate the amount of uncertainty in social media data when consumers are expressing their reaction to affective qualities. Therefore, the evaluation of user opinion or human perception requires an estimation of the amount of uncertainty. In this study, a social media analytical framework, namely SMAF, is proposed to utilise social big data to extract affective design data and evaluate affective quality along with its inherent uncertainty.

The proposed SMAF has two main components: social media data analytics and fuzzy regression analysis. The social media data analytics is developed to capture unstructured data from social media platforms which contain consumer opinion data related to affective design. The unstructured data is transformed into semi-structured data which is then transformed into numerical structured data related to the affective quality and design attributes of new products. The second component, fuzzy regression analysis, is developed to analyse affective quality from the numerical structural data generated by the first component. Although there are many machine learning technologies [83], fuzzy regression is able to predict the magnitude and the quantity uncertainty when the quantity is subjective [84], while the other machine learning approaches can only predict the magnitude. Evaluation of affective quality is subjective because it includes the evaluation of human perception. Hence, fuzzy regression techniques can be used for prediction as the relationship between affective quality and design attributes is more transparent when the fuzzy regression is used, compared to other machine learning technologies such as neural network, evolutionary computation. Therefore, the fuzzy regression method is proposed. Of all the fuzzy regression approaches [85], Urso's [86] fuzzy regression is implemented on the proposed framework since it is able to model nonlinearity while the commonly-used linear fuzzy regression cannot [87]. Also, Urso's [86] fuzzy regression is computationally simpler than fuzzy regression based on an evolutionary algorithm [88] which is computationally complex, although both methods are able to model nonlinearity.

In this chapter, section 7.2 presents a detailed background covering a wide range of literature review conducted in four major streams: (i) social media platforms adds up to success for many businesses, thus checking out social media contributions for affective design is important (ii) with various social platforms, data fusion plays an important role in ensuring information accuracy, thus we look for ways by which data fusion can be formed (iii) affective design in new product development has been reviewed extensively and, (iv) the implication of fuzzy regression techniques in affective design is also determined from past researches. Section 7.3 presents the proposed SMAF which extracts consumer information from social media and the way that fuzzy regression can be used to analyse affective quality. A case study of the affective design of automobiles is conducted to demonstrate how the proposed SMAF can be used. Section 7.4 presents the conclusion and proposed future work.

7.2 Background

A literature review of existing literature is presented in four major streams; (1) social media, (2) data fusion, (3) affective design, and (4) fuzzy regression. Section 7.2.1 summarises the research issues in general. Section 7.2.2 addresses the existing literature on social media for affective design. Section 7.2.3 focuses on the literature pertaining to data fusion, in particular that which deals with social media data and related challenges. Section 7.2.4 outlines the existing literature on affective design demand and various current approaches. Finally, section 7.2.5 focuses on the literature related to the use of fuzzy regression techniques in calculating affective quality values.

7.2.1 Outline

The rapid rise in social media use has revolutionised the power of expression in recent times. The development of the Internet and the expansion of mobile technologies have been the primary force behind the rise of social media use, providing technological platforms for information dissemination, content generation, and interactive communications. Social media use is spreading worldwide at an exponential rate. For example, in 2020, over 3.6 billion people were using social media worldwide, a number projected to increase to almost 4.41 billion by 2025 (https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/). Facebook, YouTube, Twitter, Instagram and Wikipedia are just a few of the more popular social media web applications. The amount of time people spend on these sites is also constantly increasing. For instance, Australians spend an average of 1 hour 47 minutes per day on social media, and this has been increasing slowly over the past eight years (https://www.genroe.com/blog/social-media-statistics-australia/13492). The enormous growth of social media usage has led to the generation of massive amounts of data, collectively termed ‘social media big data’ or ‘social big data’. Hence, social media landscapes can be viewed as widespread communication platforms for consumption and for sharing a rich source of day-to-day information.

Social media plays an increasingly significant role in the social lives of many individuals by offering enhanced features concerning their emotions and behaviour. For example, Facebook allows users to update their status or post information using graphic emotion icons in addition to text. Thus, different forms of data that include textual data, pictures, videos, sounds, and geolocation are provided by social media. By way of illustration – suppose an individual driving a newly-released model of a car had recently updated his views on a social media account. His feelings and behaviours toward that product can now be analysed based on his comments and the use of any emotion icons. Generally, the social media data can be divided into unstructured data and structured data, where textual content, graphic content is unstructured and followers /friends count as structured data [46]. Such information collectively constitutes online knowledge and can be analysed to obtain insights related to products, services and brands that are shared by disparate users [43]. Rathore, Ilavarasan, and Dwivedi,[43] also consider social media as a major factor influencing users’ behaviour in the form of opinions, perceptions, feedback, usage, intention, purchase habits, depth of analysis and the variety of shared information. Hence, collecting information from these interrelated multiple sources offers opportunities for more reliable and accurate knowledge which can further contribute to better business insights [44], [47].

Furthermore, social media has considerably altered the traditional way of communication and interaction between businesses and consumers. These virtual communities are now penetrating the mass public, transitioning social media communication into the main form of communication. As a consequence, social media has improved the ability for industries to integrate customers into their business model [43]. For example, in new product development, the key success factor of a new product is being able to capture the “voice of the consumers” [89]. Consequently, satisfying consumers’ requirements is not only a “need” but also a “must” for every company. Hence, it is suggested that to cover all emotional aspects of a user, a product must have these features: (1) it must be useful (it must perform the task for which it was designed); (2) it must be usable (easy to use and interact with); and (3) it must be desirable (provides feelings of pleasure and creates attraction) [90]. With social networks, designers can use online data to understand consumer requirements, affections and desirability. Hence, extracting and incorporating emotional, behavioural and demographical social media texts into the product design process may contribute to the enhancement of affective design.

Affective design involves the processes of identifying, measuring, analysing and understanding the relationship between the affective needs of the customer and the perceptual design attributes in the design domain [48]. The purpose of affective design is to satisfy the affective needs of users by integrating their affective requisites in the design attributes of a new product [49]. This gives designers the ability to generate designs which appeal more to their target market. Affective design has been shown to excite customers' psychological feelings and can help improve customer satisfaction [50]. Conventional methods of affective design focus on predicting or determining the amount of affective quality; these methods include Neural Networks [51] and the Kansei Methodology [52] which predicts consumer perceptions when the perceptual design elements of the products are given. However, perceptions are subjective; hence, uncertainties are inevitable. Therefore, the estimation of the amount of uncertainty in affective quality has received little attention in previous studies. The major works in this area include Chan et al. [49], who proposed an intelligent fuzzy regression technique to generate models for relating design variables to affective responses in which both non-linearity and fuzziness are considered. Chan and Engelke [56] presented a novel fuzzy regression method to predict affective quality and fuzziness in human assessments of specific objective features.

The aim of this study is to use fuzzy regression techniques to predict human uncertainty present in human evaluation for affective design quality. Social media platforms are a popular source of information and often contribute to the success of many businesses and organisations. However, social data does have some fuzziness present in it due to the nature of the shared information such as opinions, texts, emoticons, and comments which are nothing but human perception. Additionally, social data not only captures information from millions of individuals around the globe, but does so in real time. Hence, social media data is used to extract and analyse affective design texts and information for this framework.

7.2.2 Social Media

Social media platforms are drawing significant attention in terms of both application and research. It is not only an integral part of information ecosystems, but is also rapidly growing among users, consumers, corporations, governments and many other entities [42]. Breur [91] discusses social media as one of the four major streams of data analysis, which contains user generated data and sentiments. In addition, social media is considered a major factor influencing users' behaviour in the form of opinions, awareness, reviews, emotions, intention, purchasing habits, analysing and information sharing [43]. As a result, to leverage this, there is a need to collect and analyse social media data to extract useful patterns and investigate current trends, user affective information etc.

Several works have been submitted for social media analytics techniques, as this plays a principal role in developing tools and frameworks to collect, analyse, summarize and visualise social media data. Zeng et al. [42] discuss various challenges involved with social media analytics processes such as retrieving massive data and related metadata, computing dynamic streams of rapidly increasing data, the integration process, user generated information mining etc. In addition, Zeng et al. explore social media intelligence as a source of more productive information, although this idea seems to be in the early stages of development. Similarly, Fan and Gordon [92] discuss the various scopes of social media analytics, by means of which useful patterns and users' affective information can be retrieved. In relation to this, a framework has been presented [93] for the close analysis of affective experiences to gain marketing insights. Grassi et al. [94] created a semantic web approach to manage affective information from social media by combining semantic analysis techniques and artificial intelligence methods, though the retrieval of dynamic emotions from social media is a limitation of this approach. In order to discover these sentiments, Cambria [95] listed the major approaches for affective computing. However, these approaches did not clearly identify a common-sense knowledge base or reveal new affective knowledge in order to detect and perceive real emotions.

Another consideration is the co-creation of products through the use of social media networks, where consumers work online with company product designers to obtain user requirements [43]. It is also stated that current patterns suggest social media could produce an additional $940 billion in annual consumption, particularly in relation to the sale of electronics, hardware, software, and mobile technologies [43]. Thus, social media offers significant possibilities for product design and other commercial needs. The significance and applications of social media data have been covered extensively in previous studies. However, from a decision-making perspective, social media data still contains a level of uncertainty, which is inherent in subjective opinion based information [81].

Social media data is essentially a collection of user-generated data (blogs, posts, comments, reviews etc.) based on or influenced by personal feelings, tastes, or opinions. Hence, social data is inherently ambiguous and uncertain. Therefore, if social data is going to be utilised by companies or organizations, particularly to evaluate users' perceptions, its inherent uncertainty must be evaluated. Taking this into account, a significant need arises to estimate the amount of uncertainty in social media data when obtaining user information. In regards to dealing with the ambiguity or accuracy of data, data fusion plays an important role, as it is a way to validate information posted on multiple social platforms.

7.2.3 Data Fusion

These days, we are living in a digitized society where consumers' every single step is being recorded in some format, for which companies employ a number of social platforms in the market to stay in touch with their customers [16]. Social media data is generated from this wide range of Internet applications and web portals. Examples include, but are not limited to, Facebook, Twitter, LinkedIn and Instagram. These rapidly growing social sites allow companies to connect with users, and have created a new generation of users who are enthusiastic about interacting, sharing, and collaborating, thereby forming a new mode of communication [45]. As a result, information dissemination through social media takes place in almost every area that includes business, education, tourism, day-to-day life and health, among others. Hence, to determine the legitimacy and accuracy of the information, the data from disparate sources needs to be collated and analysed.

Information fusion ('IF'), involves the combination of information into a new set of information, aimed at reducing redundancy and uncertainty (https://en.wikipedia.org/wiki/Information_integration). Along the same lines, data fusion, which is a subset of information fusion, (or Information Integration), is the process of integrating multiple data sources to produce more consistent, accurate, and useful information than that which is provided by any individual data source (https://en.wikipedia.org/wiki/Data_fusion). It is applied in different fields where data is distributed and generated from diverse sources. Thus, to obtain a holistic view of customers, businesses ought to integrate information from multiple sources. In the field of literature related to data fusion, various methodologies have been proposed for the collection of heterogeneous information from a range of diverse sources. For instance, the crowdsourcing semantic big-data fusion approach has been used for heterogeneous media in the IoT environments, which provides higher-quality semantic fusion and more precise retrieval of information [96]. Breur [91] states that data fusion is one of four useful data analysis streams available to researchers.

Bello-Orgaz, Jung, and Camacho [45] summarise the challenges of data fusion which include: (1) obtaining more reliable methods for fusing the multiple features of multimedia objects for social media applications; and (2) studying the dynamics of individual and group behaviour, characterizing patterns of information diffusion, and identifying influential individuals within social networks.

Data fusion helps to resolve the issue of conflicting information from different sources, and identify the truth that reflects the real world. Unlike schema mapping and record linkage, data fusion is a relatively new field. It is intended to confirm the veracity of data since the web has made it easy to publish and spread false information via multiple channels [97]. To overcome the challenges of data fusion, various techniques have been developed to suit real-world applications. These specific techniques are derived from different computing areas including artificial intelligence, statistical estimation, pattern recognition, and so on [47]. However, given the very large heterogeneous dataset obtained from social media, one of the major challenges is to identify the valuable data and determine how to analyse it to develop useful knowledge [45]. Emerging big-data applications provide a solution to the integration of the heterogeneous and dynamic stream of data from diverse social media channels. The availability of consistent and trustworthy information from multiple sources is likely to provide accurate design requirements for a product, which benefits the manufacturer as well as potential customers. Data fusion can also help authenticate from social media which in turn facilitates organisational analytics. Table 7.1 shows a comparison of several research studies conducted in the field of social media analytics and data fusion, in combination with affective design.

Authors/ Title/ Year	**Social media Analytics**	**Data Fusion**	**Affective design**	
			Affective quality	Uncertainty
Grassi et al., "Semantic Web", 2011 [94]		✓		
Kwong et al., "A Modified Dynamic Evolving Neural-Fuzzy Approach to Modeling Customer Satisfaction for Affective Design", 2013 [98]			✓	✓
Lin and Wei, "A Hybrid Consumer-Oriented Model for Product Affective Design", 2016 [89]		✓	✓	

Covarrubias and Bordegoni, “Interaction with Virtual Aesthetic Shapes through a Desktop Mechatronic System”, 2014 [99]		✓		
Bello-Orgaz, Jung, and Camacho, “Social Big Data”, 2016 [45]	✓	✓	✓	
Guo, Tang, and Zhang, “CSF.”, 2017 [96]	✓	✓		
Chan et al., “A Stepwise-Based Fuzzy Regression Procedure for Developing Customer Preference Models in New Product Development”, 2015 [100]		✓	✓	✓
Zeng et al., “Social Media Analytics and Intelligence”, 2010 [42]	✓	✓	✓	
“Social Media Analytics Empowering Marketing Insight- A Framework for Analyzing Affective Experiences from Social Media Content (PDF Download Available)” [18]	✓	✓		
“A Kansei Engineering Approach to Evaluate Consumer Perception on Social Media A Case Study of Giant Manufacturing Company.pdf.” [54]	✓	✓	✓	
Rathore, Ilavarasan, and Dwivedi, “Social Media Content and Product Co-Creation”, 2016 [43]	✓	✓	✓	
Breur, “Data Analysis across Various Media”, 2011 [91]	✓	✓		
Kwong et al., “A Generalised Fuzzy Least-Squares Regression Approach to Modelling Relationships in QFD”, 2010 [101]		✓	✓	✓
Fan and Gordon, “The Power of Social Media Analytics”, 2014 [17]	✓	✓	✓	
Dong and Srivastava, “Big Data Integration.”, 2013 [97]	✓	✓		
***This thesis*, “A social media framework incorporating fuzzy regression for affective design”.**	✓	✓	✓	✓

Table 7.1 Comparison of selected papers on social media analytics, data fusion and affective design

7.2.4 Affective Design and Kansei Engineering

In today's competitive world, in order to be successful, businesses need to optimise customer satisfaction, which is essential in new product development [49]. To address customer satisfaction, Lee [102] developed a methodology to better understand user preferences based on perceived usability and perceived aesthetics. Results from this study show that before considering the actual use of a service or product, user preference was significantly affected by the aesthetics of the product rather than by usability factors. This is why many companies focus on the appearance of their products.

Affect, mood, and emotion are fundamental aspects of human beings and are found to influence people's reflexes, perceptions, cognition, social judgement, and behaviours [103]. Fong [103] presents an approach for the automatic generation of Personal Web Usage Ontology ('PWUO') of periodic access patterns from web usage logs. In this study, apart from efficiently providing users with periodic web personalisation patterns, Fong also discovered that emotional influence contributed positively to the results. Therefore, affective design plays an important role in the development of designs which better appeal to their intended market. The sole purpose of affective design is to further satisfy users' affective needs rather than solely optimizing its functional requirements. This is done by integrating the user's affective requisites within the design attributes of a new product [49]. Better affective design of a product increases that product's appeal to potential buyers and produces a more harmonious product [104].

Affective design draws its inspiration from Kansei engineering and attempts to relate subjective requirements to measurable product properties that can be tested and verified. Kansei Affective Engineering ('KE') [53] is defined as the technology of incorporating the consumer's Kansei in the product design. Nagamachi established three steps for Kansei Affective Engineering: (1) grasp the consumer's Kansei in the specific product domain using psychological or psychophysiological measurements, (2) analyse the Kansei data by statistical, medical, or engineering methods in order to clarify the Kansei structure, (3) interpret the analysed data and transfer the data to the new product domain and (4) design a new Kansei product. KE methodology has contributed to the successful development of many different products such as motor vehicles, coffee cans, beer cans, milk cartons and body cosmetics [48], [52], [54], [55].

Further, in relation to prioritising customer preferences, Chou [104] presents a Kansei evaluation approach based on the technique of computing with words ('CWW'), with the purpose of validating the classification of Kansei attributes using Kansei words, establishing priorities for customer preferences of product alternatives with respect to each attribute, and synthesising the priorities for the evaluated alternatives. Diego-Mas and Alcaide-Marzal [51] use a neural-network-based approach to present a theoretical framework which enables single-user responses to predict different product designs.

Yadav et al. [105] apply the fuzzy Kano model to quality function deployment ('QFD') with the objective of analysing the customer's aesthetic feeling toward their satisfaction. Jiang et al. [48] proposes and describes a methodology of simultaneously considering affective design and engineering specifications to determine design attribute settings and engineering requirement settings for a new product. An artificial intelligence (AI) based methodology [50], has been proposed for integrating affective design, engineering, and marketing in design specifications, at an early product design stage. This study utilises the static market trend; however, in the current technology climate, the world is evolving towards an era where online communities will define future products and services [43]. Hence, there arises a need to consider a dynamic market in relation to affective design. In addition, inappropriate affective designs can only be determined by past affective information and data. Hence, there is a need for an approach that continuously updates users' affective quality information.

Much previous research has been conducted to estimate affective quality for better affective design. However, data relating to affective satisfaction is inherently vague or uncertain. The above-mentioned approaches cannot adequately address this issue. In order to estimate the level of uncertainty or vagueness, recent research has shown that the fuzzy regression model is a more commonly-used method for developing consumer preference models. The fuzzy regression model is explicit, meaning that analytical information can be identified for use in new product development, and developed models can address the fuzziness in consumer preferences [100]. Chan and Engelke [56] propose a novel fuzzy regression method to predict affective quality and estimate fuzziness (i.e. vagueness or uncertainty) in human assessments, when objective features are given. So far, studies have targeted fuzziness in observed survey data or static data, yet there seems to be no research on estimating the effect of fuzziness on affective design datasets, which varies with time and can be captured online through the use of social media. Table 7.2 gives a summary and comparison of the literature conducted in the field of data fusion, fuzzy regression, and perceptual uncertainty in conjunction with affective design.

Authors/ Title/ Year	Data Fusion using social media analytics	Fuzzy regression	Affective quality	Perceptual uncertainty
"Affective Design with Kansei Mining An Empirical Study from Automotive Industry in Indonesia.pdf." [106]			✓	
Shen and Wang, "Affective Product Form Design Using Fuzzy Kansei Engineering and Creativity", 2016 [107]			✓	
Yadav et al., "Prioritization of Aesthetic Attributes of Car Profile", 2013 [105]			✓	
Ohbyung Kwon, Choong-Ryuhn Kim, and Gimun Kim, "Factors Affecting the Intensity of Emotional Expressions in Mobile Communications", 2013 [108]	✓			
Chan and Engelke, "Varying Spread Fuzzy Regression for Affective Quality Estimation", 2017 [56]		✓	✓	✓
Kao et al., "Co-Creating Value with Consumers through Social Media", 2016 [109]	✓		✓	
Kwong, Jiang, and Luo, "AI-Based Methodology of Integrating Affective Design, Engineering, and Marketing for Defining Design Specifications of New Products", 2016 [50]		✓	✓	
"Zhang - The Role of Perceived Affective Quality.pdf.", 2004			✓	

Kwong et al., “The Hybrid Fuzzy Least-Squares Regression Approach to Modeling Manufacturing Processes”, 2008 [110]		✓	✓	✓
Chan et al., “An Intelligent Fuzzy Regression Approach for Affective Product Design That Captures Nonlinearity and Fuzziness”, 2011 [49]		✓	✓	✓
Cho and Kim, “A Study on Affective Design by Subjective-Objective Co-Approach”, 2014 [111]			✓	
“Decision Support for the Design of Affective Products: Journal of Engineering Design: Vol 20, No 5.” [52]			✓	
Chan et al., “A Flexible Fuzzy Regression Method for Addressing Nonlinear Uncertainty on Aesthetic Quality Assessments”, 2017 [112]		✓	✓	✓
***This thesis*, “A social media framework incorporating fuzzy regression for affective design”**	✓	✓	✓	✓

Table 7.2 Comparison of selected research papers on data fusion, fuzzy regression, affective quality and perceptual uncertainty

7.2.5 Fuzzy Regression

Previous studies have acknowledged the significance of good affective design in new product development. In order to evaluate affective design, researchers have introduced different methodologies to measure the amount affective quality/aesthetic quality by either ‘surveying in order to obtain users’ subjective perceptions of object aesthetics’ or by ‘predicting affective qualities by means of objective features based on product design attributes’ [112]. A number of studies used survey techniques. For example, Diego-Mas and Alcaide-Marzal [51] used a neural-network-based mathematical model to present a theoretical framework which enables single-user responses to predict different product designs. This is done using limited survey responses and thus has less statistical power. Similarly, a large amount of research has been conducted using the Kansei engineering methodology [43, 54, 111] to better understand consumer perception and affective requirements. These studies were conducted using surveying techniques. However, because many of the survey questionnaires contain too many subjective evaluation items, there is a possibility for uncertainty to be present in the subjective evaluations given by participants. However, the uncertainty of survey participants in answering the survey data may not be analysed using traditional statistical methods. In addition, it is infeasible to conduct a survey for every single design attribute of products. Surveying is time consuming and limited features or questionaries', with limited access to the population of concern [112].

Another approach for predicting affective quality is based on the assumption that the objective features of products are correlated with the user's perception of affective quality [56]. As a result, empirical models can be built using objective features such as colours, product style, outlook, interfaces etc. in order to determine the affective quality of a product. Statistical regression techniques are generally used to estimate the correlation and significance of variables [56, 112]; however, the uncertainty in subjective human evaluation cannot be taken into account. Since human emotions are uncertain, crisp values from statistical regression do not correspond effectively when assessing affective quality. To overcome this, the fuzzy regression methodology was developed, which evaluates the uncertainty in human perceptions.

A number of studies have used fuzzy regression techniques to estimate affective quality and uncertainty in human perception [49, 56, 104, 105]. In this context, products with strong affective quality are easy to evaluate. However, products with relatively moderate affective quality have a significant level of uncertainty.

To illustrate the uncertainty of affective quality, an automobile case study is shown in fig 7.1

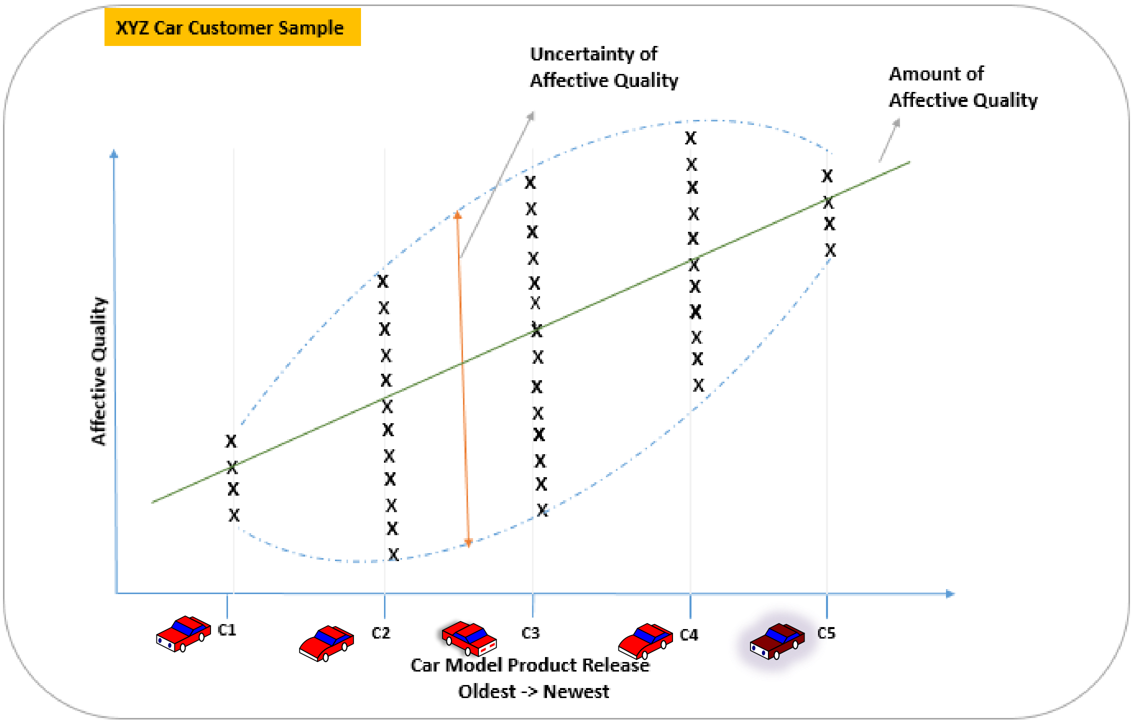

Figure 7.1 Uncertainty in affective design

For instance, suppose C1, C2, C3, C4, and C5 are different car models. As shown in Figure 7.1, C5 is the most recent model with the highest price and the best affective design when compared to C4. C4 is the second most recent model with better affective quality than C3. C3 is the third most recent model with better affective quality than C2, and so on. Likewise, C1 is the basic model vehicle with the lowest price and the lowest affective design. As the affective quality evaluation is subjective, users may have more confidence in determining the affective quality of C5 and C1. Hence, the evaluation of uncertainty is low. On the other hand, for middle-range models such as C2, C3, and C4, users have less confidence in evaluating the affective quality. Therefore, the uncertainty in evaluating the affective quality is higher for the average model vehicles, which have average affective qualities. Although Chan and Engelke [56] have developed a fuzzy regression method for predicting affective quality and uncertainty when evaluating affective quality, fuzzy regression coefficients are determined based on heuristic algorithms, which are time consuming and indeterminate. A more time-effective and determinate approach is essential to determine fuzzy regression coefficients. This approach will determine the affective design of a product (e.g., car model) by taking design attributes of a car such as body type, colour, transmission etc. as independent variables to predict dependent variables, i.e., magnitude of affective quality values and uncertainty in evaluating affective quality which will help to achieve the affective design of a car.

7.3 Summary of Research Issues

The comprehensive study of the literature in the areas of social media analytics, data fusion, affective design and fuzzy regression reveals that in the current dynamic world of the internet boom, manufacturing companies need to be aware of consumer's affective needs and perceptions when marketing or designing their products. Those affective and perceptive needs can be effectively captured from social media channels. Although a lot of work has been done in assessing affective design attributes, affective quality and most preferred design values, there has been limited focus on the area of determining perceptual uncertainty in affective quality evaluation. The below research concerns have been identified from the literature review

7.3.1 In the area of affective design, there is a lack of data fusion applied to multiple social media channels, particularly concerning new product development, although data fusion has been applied to many areas such as business intelligence, business performance management and bio-informatics.

7.3.2 There is a lack of research focusing on the estimation of uncertainty in affective quality evaluation using social big data and fuzzy regression techniques. Most research has focused on determining the magnitudes of affective qualities by consuming limited survey data.

This framework will produce data fusion for affective design using social big data and will incorporate a fuzzy regression approach, to determining perceptual uncertainty while evaluating affective quality in new products.

7.4 Methodology

The proposed methodology addresses affective design in new product development while implementing fuzzy regression techniques.

A new social media analytics framework (SMAF) is developed to extract social media data from Twitter and Instagram. Two fuzzy regression techniques are applied to evaluate the affective design quality and uncertainty in the perception evaluation.

The method is demonstrated by using a case study involving the development of an automobile design. In general, the design phase includes steps such as the prediction of future trends and changing customer needs, clay and digital modelling and designing of interiors, all of which are combined to form a new model. This study considers a luxury family car and examines the affective design perceptions along with the evaluation of uncertainty present in the design perceptions in order to arrive at a precise evaluation.

7.4.1 Social media analytical framework

The proposed Social Media Analytical framework (SMAF) is illustrated in Figure 7.2. SMAF has two main components: social media data analytics and fuzzy regression analysis. The first component, social media data analytics, involves data extraction, data acquisition, data storage and data preparation followed by data analysis. These phases are intended to capture customer opinions from the social media. They also extract the design attributes and affective quality from the customer opinions. The second component, fuzzy regression analysis, is used to analyse the magnitude of affective quality and uncertainty when the consumers are evaluating the affective qualities.

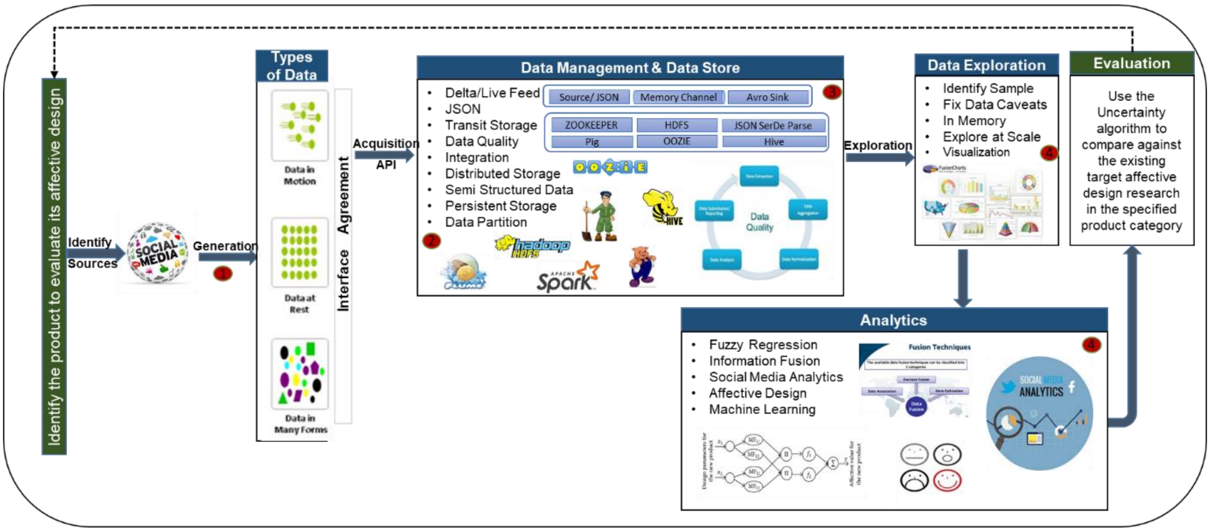

Figure 7.2 Social Media Analytical Framework (SMAF)

Social Media Analytical Framework consists of four phases to create design database. The four phases cover the big data life cycle suggested by Hu et al.

The four main phases are

1. Data extraction

Extracting information from social media channels generates heterogeneous types of unstructured data and requires a compatible Application Programming Interface (API), which is a software program that has a specific way of accessing and retrieving data. Generally, social sites have their own compatible APIs for connectivity.

The Twitter and Instagram social media platforms are utilized in this framework. For Twitter, REST API 2.0 was used to extract data from Firehose, whereas Instagram's posts are extracted manually. For Twitter, an automated PHP program was written to call the Twitter API which extracts data based on keywords in JSON format. The aim of this part of the study is to evaluate the affective design of luxury family car models. A key-word based search yielded several models in the luxury car category: BMW X5, Mercedes GLE, Land Rover sport, Audi Q7, Lexus RX350 and Porsche Cayenne. Then, the PHP program is applied to extract tweets related to the design of luxury car models. Keywords such as "BMW X5", "BMW X5 colour", "Mercedes GLE colour", "AUDI Q7 rooftop", "Lexus bumper" were used to extract data. About 100 to 200 tweets were extracted.

2. Data Acquisition/transformation

The extracted original twitter data is in JSON form with complex multilevel nesting, with as many as 30 levels. Due to the complex nesting structure, it is difficult to convert data into a structured form by means of online tools. Therefore, SAS functions and Procs are used to first convert the JSON data into semi-structured data, i.e., with column, value combinations. Table 7.3 shows semi-structured data sample which has three columns. The column details are as follows

Record_Number: Represent tweet number. For instance, if a single tweet has 75 attributes, then there will be a row for each i.e., 75 rows in total, thus a single tweet can have 70 - 300 rows. And Record_Number against all these 75 rows identifies the unique tweet.

- Column_Name: This contains the name of the column. Tweeted data has multi-level hierarchy, so the column name is a combination of all hierarchy. For example-
- Metadataresult_type: in column 3: Under "Metadata" what is the "result_type" of tweet. This is two level hierarchy: 1st Parent: Metadata and, 2nd Child: result_type.
- retweeted_statusmetadataresult_type: Starting from prefix, "retweeted_status" means that the tweet is retweeted, and this hierarchy will only appear if the tweet is retweeted. Under retweeted_status, what is the "metadata" of the retweet, and within this the particular metadata information looking for is the "result_type" which could be "recent", "mixed" and "Popular". (Where, mixed: Include both popular and real time results in the response; recent: return only the most recent results in the response; popular: return only the most popular results in the response).

- retweeted_statususerdescription: The status description of user retweet (3 hierarchy/level - super parent - retweeted_status, parent - user, child - description)

Value: This column contains the defined values.

This semi-structured data is converted into a tubular CSV format of which JSON is transformed to CSV by the conversion tool.

record_number	value1	column_name1
1	"recent"	�etadata�esult_type
1	"en"�	�etadata�so_language_code
1	"Wed Feb 12 05:33:02 +0000 2014"	�reated_at
1	4.33E+17	�d
1	"433473801863856128"	�d_str
1	"Hypertension Strengths and limitations of the JNC 8	�ext
1	"<a href=\"http:\/\/www.hootsuite.com\" rel=\"nofo	�ource
1	FALSE	�runcated
1	null	�n_reply_to_status_id
1	null	�n_reply_to_status_id_str
1	null	�n_reply_to_user_id
1	null	�n_reply_to_user_id_str
1	null	�n_reply_to_screen_name
1	348349430	�ser�d
1	"348349430"	�ser�d_str
1	"Blood Bio"	�ser�ame
1	"blood_bio"	�ser�creen_name
1	""	�ser�ocation
1	"BioPortfolio's Blood newsviews�description": "BioPo	�ser�escription
1	"http:\/\/t.co\/M2zMi3VX2d"	�ser�rl
1	"http:\/\/t.co\/M2zMi3VX2d"	�ser�ntities�rl�rls�rl
1	"http:\/\/www.bioportfolio.com\/channels\/blood"	�ser�ntities�rl�rls�xpanded_url
1	"bioportfolio.com\/channels\/blood"	�ser�ntities�rl�rls�isplay_url
1	[0	�ser�ntities�rl�rls�ndices
1		�ser�ntities�rl�rls�

Table 7.3 Semi-structured data sample

3. Data storage and data preparation:

MS Excel and SQL enable the storage and querying of the database. Tweets are analysed with a combination of automated and manual processes. Firstly, an automated code is developed using Python and NLTK to analyse the sentiment keywords such as ‘good’, ‘bad’ and ‘neutral’ for design quality. Later, the affective design database is examined extensively to extract useful and relevant texts for analysis purposes and to discard tweets that are irrelevant. Extracted tweet texts contain data about design attributes such as colour of car, model, trunk, hood, kidney grille and rooftop of a car. Missing values and null values are also removed from the database. In total, 150 affective design tweets are selected to be stored as “affective design database”.

4. Data analysis

To understand users' affective design needs, the database is extensively analysed for different design attributes. Colour, sporty, muscle car, quietest, supercar, dream car, roofless, and kidney grille are the major affective design attributes found from the social media database for family luxury car models. The affective design attributes are scaled into categorized/continuous variables i.e.

i. Colour
ii. Car types (such as sedan, hatchback, SUV)
iii. Rooftop (such as sunroof, full sunroof, wide sunroof, no sunroof)
iv. Shape (such as sporty, muscular)
v. Kidney grille (such as large, medium, small)
vi. Overall outlook (such as supercar, dream car, fantastic, quietest)

7.4.2 Fuzzy regression techniques for affective quality evaluation

Affective quality evaluation of products is subjective since it involves human perception. Therefore, affective qualities involve a certain amount of fuzziness or uncertainty. To analyse the affective quality of a product, both the amount of affective quality and the uncertainty inherent in the evaluation of affective quality must be predicted. The affective quality of a product can be predicted by the model $f(.)$ in (1):

$$\tilde{y} = f(\bar{x}) \qquad (1)$$

where $\tilde{y} = (y_c, y_s)$ is the affective quality in a fuzzy number, y_c is the magnitude of affective quality and y_s is the uncertainty when evaluating the affective quality; $\bar{x} = (x_0, x_1, ..., x_n)$ is the set of design attributes; x_i is the i^{th} design attribute of the product. The original fuzzy linear regression model was first developed by Tanaka et al. [113] with the following proposition:

Given the fuzzy parameters $A_i = \{\alpha_i, c_i\}$, with $i = 0,1,2,...,n$, the fuzzy linear function can be formulated as

$$\begin{aligned}\tilde{y} &= A_0 + A_1 x_1 + ... + A_n x_n \\ &= (\alpha_0, c_0) + (\alpha_1, c_1) x_1 + ... + (\alpha_n, c_n) x_n , \\ &= \bar{\alpha} \cdot \bar{x}^t + \bar{c} \cdot \bar{x}^t\end{aligned} \qquad (2)$$

where $\bar{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_n)$ and $\bar{c} = (c_0, c_1, ..., c_n)$ are the fuzzy parameter sets. The membership function of $\tilde{y}$ is given as:

$$\mu_y(y) = \begin{cases} 1 - \dfrac{|y - x^t\alpha|}{c^t |x|}, x \neq 0 \\ 1, x = 0, y = 0 \\ 0, x = 0, y \neq 0 \end{cases} , \quad (3)$$

Where $|\bar{x}| = (|x_1|, ..., |x_n|)$. In equation (3), $\mu_y(y) = 0$ when $\bar{c}^t |\bar{x}| \leq |y - \bar{x}^t \cdot \bar{\alpha}|$.

Equation (2) shows that only the linear terms exist, and no interaction or high order term is present. Therefore, the nonlinear relationship between $\bar{x}$ and y cannot be taken into account. The nonlinearity between $\bar{x}$ and y cannot be modelled when Tanaka's fuzzy regression is used, which is a limitation of the Tanaka approach.

Urso's fuzzy regressive method [86] can also be used to study the relationship between the affective quality and design attributes of the product. This approach attempts to overcome the limitation of Tanaka's fuzzy regression which cannot model the nonlinearity. In Urso's fuzzy regression, a fuzzy number in the triangular case is considered as $F = (c, p, q)_{LR}$, where *c* is the centre which describes the quantity; *p* and *q* are the left spread and right spread which describe the uncertainty; *L R* denotes the left and right shape function. A polynomial regression with fuzzy data is defined to analyse the relationship between the prediction of the fuzzy quality $\tilde{y} = (c, p, q)$ and the design attributes x_i with $i = 1, 2, ..., m$:

The centre of the fuzzy quality:

$$c_i = a_0 + a_1 x_i + a_2 x_i^2 + ... + a_m x_i^m + \varepsilon_i \quad (4)$$

The left spread of the fuzzy quality:

$$p_i = d + b(a_0 + a_1 x_i + a_2 x_i^2 + ... + a_m x_i^m) + \lambda_i \quad (5)$$

The right spread of the fuzzy quality:

$$q_i = h + g(a_0 + a_1 x_i + a_2 x_i^2 + ... + a_m x_i^m) + \rho_i \quad (6)$$

Where *m* denotes the highest order of the design attributes of the fuzzy regression; all a_i with $i = 1, 2, ..., m$, b, d, h, and g are the fuzzy regression coefficients; λ_i, ρ_i, and ε_i are the residuals.

Tanaka’s linear model [113] shows that errors in the evaluation of fuzzy data are caused by the observational errors as well as the fuzziness when consumers are evaluating the affective quality. In Tanaka’s linear model, the quality evaluation can be explained by the fuzzy spreads. Urso et al.’s model [86] in (2) to (4) presented a polynomial regression which can be used to address the nonlinearity between affective quality and design attributes since the higher order terms exist in the Urso et al.’s model in (4) to (6).

The predicted uncertainty is indicated by the left spread in (5) and the right spread in (6). This means that greater spreads have high uncertainty and smaller spreads have low uncertainty. Since Urso’s fuzzy regression can be used to model the nonlinearity between the affective quality and design attributes, the Urso’s fuzzy regression are proposed to integrate with the proposed Social Media Analytical Framework (SMAF) which is discussed in previous section. The details regarding the determination of the coefficients of (4) to (6) can be found in [86]. The proposed framework can be used to analyse affective quality and uncertainty for evaluating the affective quality, when the consumer reviews and opinions for a product are available on the social media channels such as Twitter, Instagram and Facebook.

7.4.3 Evaluation and analysis using a case study

Generally, when consumers choose a new car model, colour is considered an important affective design factor. The colour of a car can influence the purchase as much as its features and functionality can. Hence, a case study is discussed here utilizing SMAF method to predict the uncertainty present in affective perceptions of design quality. The data has been extracted and analysed using SMAF; this is followed by an evaluation using fuzzy regression techniques.

7.4.3.1 Case Study

To determine the uncertainty present in users’ perception of design, a prediction model is proposed. The fuzzy regression technique is utilized in this study to develop a prediction model to predict fuzziness or uncertainty present in social media design texts. In fuzzy regression, numerical parameters are desirable. However, design features are usually categorical variables. For example, the body type of a car can be categorised as sedan, hatchback, coupé, van, compact or SUV. Hence, of the various design features, colour is identified as both a categorical and continuous variable. The reason for continuous is that colours are associated with the wavelength of the light that is being reflected by objects. Therefore, colours can be classified as continuous using different wavelength values of the visible light spectrum (red, orange, yellow, green, cyan, blue and purple) as shown in Figure 7.3. Because black is known as the absence of any colour, it has the value of "0" in the wavelength. Conversely, white occurs when the wavelengths of all colours are combined. Therefore, white is situated in the middle of the colour range, close to green. In fuzzy regression, the design attribute of colour is considered as crisp data to predict the magnitude of affective quality and the uncertainty in determining the quality of a car based on colour.

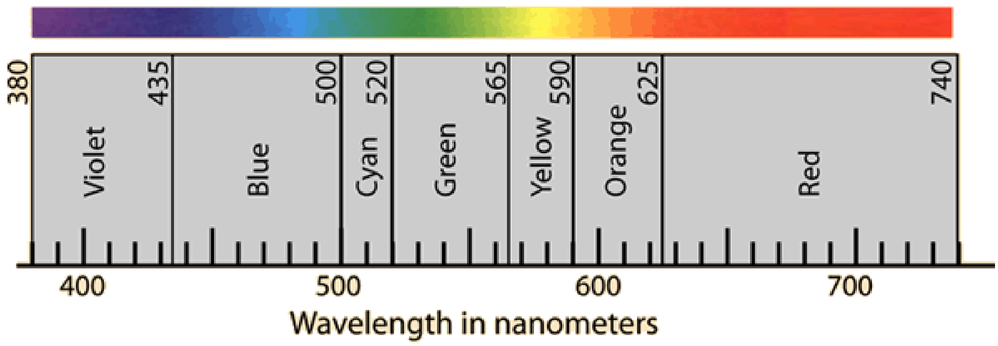

Figure 7.3 Colour Wavelength

In order to perform this case study, colour tweets from Twitter design database discussed in step 4 (data analysis section) has been searched and analysed. Thirty colour-related tweets associated with car are selected for the case study. A screenshot of several tweets is shown in Figure 7.4.

Colour	Tweet
red	@BMW @BMW_SA Concept 4 showing what BMW's will look like in near future, love the colour too! #TuesdayMorning #TuesdayThoughts #TuesdayMotivation #BMWConcept4 #BMW https://t.co/adVmn9g8K4 https://t.co/SHhDZKU6ln
red	#RT @CarsSouthAfrica: RT @HannesCars: The Porsche Taycan in red. Not so sure the colour works on it. White is much better. #IAA2019 #iaa #FrankfurtMotorShow @CarsSouthAfrica https://t.co/gY6PThXuvW
namiborange	RT @carspecialists: All the staff are loving the #colour of this stunning 2017 Land Rover #Discovery 3.0L TD6 HSE with Dynamic Pack. ...it's called #NamibOrange if you're wondering.👌 #LandRover #landroverdiscovery #LandroverLove #Sheffield #adventure #explore #countryside #Derbyshire https://t.co/6I2tLgbcPz
green	@dammy_porsche Lol ***green colour***
green	RT @DRIVETRIBE: Only haters will dislike the colour 👀 https://t.co/prNbGp8ull+
atlantis blue	Atlantis Blue Might Just Be The Best Colour In #BMW's Palette https://t.co/duH8B6mNTk via @Carscoop https://t.co/ucI7b3Cp0C
blue	RT @RiskyPleasures: #Porsche I usually like the black matte rims, but this colour which I LOVE and these rims are a perfect pair! https://t.co/hsfODvBbk9
blue	RT @CarsSouthAfrica: Fresh from the #FrankfurtMotorShow... Shots of the new @BMW_SA M135i in an interesting colour. #IAA2019 #iaa https://t.co/CPUuJ2RNWr
navy blue	RT @equatormotorsuk: I like the colour on this Mercedes Benz G63 😍 https://t.co/iWGj1qbjNB via @godsxn
vantablack	@rvcingseb The only colour an X6 should be: https://t.co/CfO84tCaKH So you can't see it 😂

Figure 7.4 Sample design tweet screenshot

A tweet is analysed by means of a five-point scale to illustrate the design quality magnitude and uncertainty. "5" is the highest affective quality, followed by "4" as good, "3" as neutral, "2" as bad and "1" as lowest. Similarly, uncertainty is measured on a five-point scale, where the range 1 and 2 is considered as "low", 3 as "medium" and 4 and 5 as "high". Table II shows a five-point scale which illustrates uncertainty and affective quality in the data related to colours.

7.4.3.2 Evaluation using Fuzzy regression

Urso's fuzzy regression [86] predicts centres and spreads and checks their interdependence. The model in the polynomial regression form is more appropriate for the samples than is Tanaka's fuzzy regression which only has linear terms. Urso's fuzzy regression is coded and executed on MATLAB. Figure 7.5 shows the affective quality, y, vs colour wavelength, x, for the fuzzy regression model. The affective quality for the order three fuzzy regression model is shown in Figure 7.6. To interpret results from the graph; blue lines shows affective quality magnitudes. Results of uncertainty are shown by the two boundaries of the red dotted lines in the graphs. The samples are illustrated by the '+'. The results show that the uncertainty is low for medium wavelength colours such as "White", whereas the uncertainty is high for a low wavelength such as "Black". Also, the uncertainty is high for high wavelength such as "Red".

Table 7.4 Five pointer scale illustrating affective quality and uncertainty in design tweets

Design Tweets	**Five-point scale**		
	Sentiment	***Uncertainty (ys)***	***Affective quality (yc)***
"I've just returned from a few days in Istanbul, on my last day I saw a Porsche Boxster in metallic brown with beige leather interior!? I can't imagine seeing one that colour in London, and it wouldn't be my choice of colour scheme, but hey ho, it looked errrrr individual □ https://t.co/eQOu3yEFJH"	Very bad	Low(1)	1
"@dammy_porsche Lol green colour"	Bad	High(3)	2
"The X6 Vantablack features the world's blackest black paintjob, a colour which was introduced in 2014 by a brand known as Surrey NanoSystems. What do you have to say about the blackest black? https://t.co/h5PYO0bf4H"	Neutral	Medium(2)	3
"RT @equatormotorsuk: I like the colour on this Mercedes Benz G63 ☺ https://t.co/iWGj1qbjNB via @godsxn"	Good	Medium(2)	4
"Atlantis Blue Might Just Be The Best Colour In #BMW's Palette https://t.co/duH8B6mNTk via @Carscoop https://t.co/ucI7b3Cp0C"	Very good	Low(1)	5

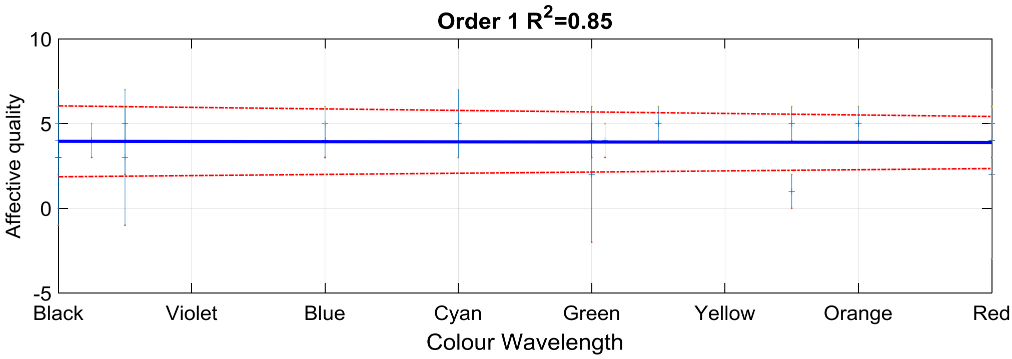

Figure 7.5 Order 1 fuzzy regression model

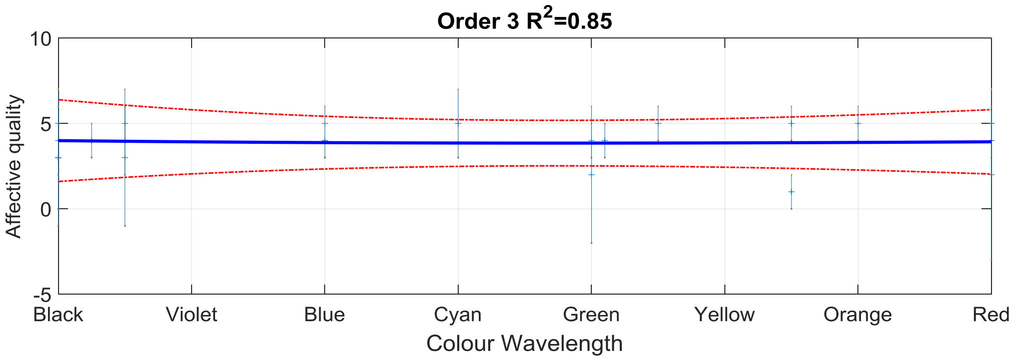

Figure 7.6 Order 3 fuzzy regression model

Table 7.5 Affective quality prediction

Fuzzy regression models		Affective quality Prediction ($\tilde{y}$)
Order 1 Fuzzy regression model	Centre	$c = 3.9517 - 0.0100 \times x$
	Left spread	$p = -29.4258 + 7.9760(3.9517 - 0.0100 \times x)$
	Right spread	$q = -29.4258 + 7.9760(3.9517 - 0.0100 \times x)$
Order3 Fuzzy regression model	Centre	$c = 3.9965 - 0.0845 \times x + 0.0138 \times x^2 - 0.0005 \times x^3$
	Left spread	$p = -26.6181 + 7.2597 \times (3.9965 - 0.0845 \times x + 0.0138 \times x^2 - 0.0005 \times x^3)$
	Right spread	$q = -26.6181 + 7.2597 \times (3.9965 - 0.0845 \times x + 0.0138 \times x^2 - 0.0005 \times x^3)$

7.4.3.3 Results and discussion

Figure 7.6 shows that there is a better fit when the order 3 model is used. The figure illustrates more clearly that the uncertainty of medium wavelength is low, whereas the lowest wavelength has high uncertainty. Both figures shows that the affective qualities are almost the same when the different colours are used. However, Figure 7.6 shows that the uncertainty of consumer evaluation is smaller when the colours of cyan, white, silver and green are used for the car design. This information provides insights for automobile companies that want to design and manufacture new cars to satisfy consumer affective needs related to colour. Based on this information, automobile companies can better design a car such that the affective quality is high and uncertainty is low. The order 1 model in Figure 7.5 and the order 3 model in Figure 7.6 are given in Table 7.5. The models of fuzzy quality $\tilde{y}$=(c,p,q) in the form of equation (4), (5) and (6) are shown in the table. The two models show that the spreads are symmetrical which is same as the characteristics shown in Figure 7.5 and Figure 7.6. The “order 3 model” illustrated in Figure 7.6 is better fitted, compared to the “order 1 model” in Figure 7.5. The characteristics of wavelengths and the uncertainties indicated by the spreads are illustrated more clearly in the figure. The uncertainty of medium length is low, whereas the lowest and highest wavelengths have higher uncertainties.

7.5 Conclusion

The paper proposed a social media analytics framework (SMAF) to predict uncertainty in affective design integrating fuzzy regression. The SMAF is used to extract and analyse affective design data from social media platforms such as Twitter and Instagram. Affective quality perceptions are evaluated using the fuzzy regression. A case study of the affective automobile design was used to demonstrate how the SMAF is used. The colour feature of a car is used as the design attribute to predict the uncertainty present in the affective quality of car models. Results shows that affective quality is almost the same for all colours. However, uncertainty is low for medium wavelength colours such as white and high uncertainty exists for extreme wavelengths. Thus, uncertainty is high for the colours with lowest and highest wavelengths, i.e., the black and red car models. Based on the predicted uncertainty, automobile companies obtain more insight so that they can manufacture new cars to suit the markets. The study can be enhanced by considering multiple design attributes or using categorical design features.

Chapter 8:

8 Big data and Machine Learning use in healthcare industry

This chapter provides

- An overview of health industry and big data.
- Two health industry projects, undertaken as part of the doctoral internship are discussed
- The internship project problems and approaches has been presented in the form of case studies.
- A new roadmap is developed as part of the thesis contribution to assist in developing prediction tool for health industry.

8.1 Introduction

With the adoption of electronic medical systems, health-related data such as Electronic Health Record (EMR/EHR), patients hospital information, IoT, medical claims and images have become readily available which results in big data in the healthcare domain [57]. In healthcare, any device that generates data about a person's health and sends that data into the cloud will be part of the IoT. Wearables such as smart watch and devices that track the number of steps taken by wearers, their heartrate, their weight, and the trend are perhaps the most familiar example of such devices. Hence, there are various sources of big health data that include hospital records, medical records of patients, results of medical examinations, and devices that are a part of Internet of things [114, 115]. The significant potential for the application of big data has already emerged in the healthcare industry with the use of predictive analytics, devices and systems used by healthcare payers for fraud prevention, to determine patients who require a management care plan, and to identify and track diabetes type II patients. Big data is also offering tremendous opportunities for evaluating and analysing health care quality [116]. Most health systems have a great number of opportunities to improve clinical quality and financial performance, and analytics are required to identify and take advantage of those opportunities. However, there are several obstacles that need to be overcome in relation to big data: siloed data, fragmentation, privacy, high costs, and questions about data ownership (big data in digital healthcare).

Machine Learning (ML) provides effective use of big data in healthcare by enabling the development and deployment of predictive modelling and analytical approaches. ML approaches are often interchangeably used with artificial intelligence (AI) approaches. ML and AI now make it possible to unravel the patterns, associations, correlations and causations in complex, unstructured, non-normalized, and unscaled datasets that the big data era has introduced (Camacho et al. 2018). Therefore, the use of ML is increasing in the medical/healthcare sector as it is in other sectors. The incorporation of ML in the health industry is a popular strategy applied to improve health care delivery [117]. Health care providers, private health insurance companies and hospitals are all focusing on the application of ML and AI techniques to solve complex problems, using patient database and related information to create prediction models. ML assists with the development of tools for medical decision-making, thus it may help patients to make decisions without consulting physicians in future, according to researchers [117]. ML models have already performed better than other models in, for example, predicting risk of readmission to the hospital emergency unit, detecting health risks such as diabetes type II, cardiology problems, drug efficacy, algorithm for radiology etc.[58, 59, 61]. However, ML models in healthcare do have many methodological biases such as dataset biases, small sample size, confounding variables and clinical biases such as practical clinical assessment of diagnostic or prediction models, and ethical concerns which have been recently been the focus of several researches [118, 119]. Having said that, ML helps in developing a fair prediction model which helps to remove human racial biases such as white/black skin, gender biases and so on, which have been ongoing social issues. Also, it reveals any hidden patterns in the data which can offer new insights or indicate statistically strong relationships among variables for analysis that may or may not be present in the current health industry. This study extensively analyses ML aspects as part of the work done during the doctoral internship in the health industry.

8.2 Literature Review

In recent years, the development of computer software and hardware technologies in the health industry have enabled the generation of massive amounts of data. In addition, with the adoption of electronic medical systems, health related data such as Electronic Health Record (EMR/EHR), medical claims and images have become readily available, generating additional big data [57]. This "big health data" has huge potential to assist in healthcare decisions or treatments [57-60]. However, it is evident now that big health data is useless unless it is analysed or interpreted correctly to identify useful patterns or information [61]. When applying ML and AI techniques to big health data, algorithms that could assist with medical treatment or diagnosis, and provide management care can be determined out of datasets and transformed into tools or acted upon as a service. Thus, ML has introduced advanced healthcare options and has helped to solve complex health problems. It is also expected that AI tools for the health industry will have a $20 billion market within next five years [118].

ML works wonders for the development of models while providing generalized algorithms derived from the large volume of data. An effective model can make a whole sample dataset or massive big health data understandable in the same way that simple mathematical equations are capable of explaining physics, according to Google's chief scientist [120]. However, it is not always easy to create a good model for the health care industry for various reasons such as lack of expertise, security issues such as the privacy policy of patient data, data ownership issues and the variety of data such as images, diagnosis/treatment information for patients, discharge reports or summary text reports which require translation and data linkage activities before the data is analysed. Health data is siloed and fragmented which either requires many layers of data access or results in missing fragments of data.

Regardless of the aforementioned complexities of health data, ML can avoid biases in the treatment and diagnosis of the patients. Generally, there are social biases related to race, gender or ethnicity. This issue is solved to a certain extent by the application of ML, as the models are an output of significant important variables which are chosen by ML but not by humans [119].

ML utilises supervised and unsupervised methods to create models. Supervised methods include the task of inferring a function from labelled training dataset. By fitting to the labelled training set, the idea is to find the most ideal model parameters to predict unknown labels for other objects (test set) (https://blog.statsbot.co/machine-learning-algorithms-183cc73197c). If the label is a real number, the algorithm is known as regression. If the label is from a limited number of values, where these values are unordered, then it is known as classification. This study adopts ML supervised algorithms where the target variable 'y' is produced using predictor variables 'x'. Generally, a machine learning model requires large amount of data for training. The models are trained to achieve high accuracy; however, more recently, there has been an increasing demand to understand the way a specific model operates and comprehend the underlying reasons for the decisions made by the model in medical science. It is crucial to understand the story/insights of the model in the area of medicine/ health to improve health services and add new opportunities. Hence, an explanation of a machine learning model developed for the health problem should be explained well to ensure the importance of model outcome and its limitation [121].

Various computational models for the health industry have been developed using ML approaches [58, 122, 123]. For instance, the LACE model uses four variables to predict risk of readmission (derivation and validation of an index for death). Prediction of admission to the hospital emergency unit has also been piloted using ML techniques [60]. Supervised ML regression is used to develop a model to predict patients' risk of contracting a multidrug-resistant urinary tract infection (MDR UTI) after hospitalization [124]. Probabilistic Neural Network and Random Forest are used to develop an ML pre-diagnosis model for the emergency department utilising patients' verbal complaints [59]. Although ML is becoming established in the health care domain, there are discussions about the fairness of the model, its ethical and methodological limitations, and a clear explanation and correct interpretation of models for correct use [118, 119, 125, 126]. The main focus of this study is to present work done as part of a doctoral internship undertaken in health care industry, followed by discussion and a road map which will assist in building a prediction model. The first internship involved machine learning being used to build a novel prediction model for an Australian Health Insurance Company. As part of another engagement with an Australian public hospital's innovations project, the optimisation process for a prediction tool is discussed with a roadmap to optimise the model. Furthermore, the limitations of machine learning are discussed and explored with a roadmap and in order to understand the application of ML aspects for the health industry.

8.3 Organizational research and innovation data science projects

Big data and machine learning techniques are evolving in the health industry and are now used to develop tools and automated systems to predict diagnosis, treatments, risks and services for better support and quick decisions. Along the same lines, one of the Australian private health insurance providers is trying to understand customers' needs by means of an analytics approach, which was used for the internship research project. A solution is build a model using a predictive modelling approach of machine learning, the objective of which is to predict the main health conditions of specific customer segments, which are then used as a basis for the development of new products. The models and ML approaches are demonstrated by means of a case study; this is followed by discussion of the various features and limitations of ML.

In addition, the innovative project of an Australian public hospital is also presented as a case study where machine learning approaches are used to optimise the existing prediction tool. A new model is presented for readmission prediction, which is discussed along with the limitations and suggestions for dealing with the limitations. Finally, a roadmap is developed for the health industry, which was devised during the research work and the internship.

Section 8.3.1 presents case study 1, where a novel model is developed to predict key health conditions based on clients' lifestyles and historic medical data. Additionally, a predictive model is built to predict the health services pattern for different client segments. The research project is initiated by an Australian private health insurance firm with an objective to identify health needs and health services usage pattern for their clients. ML is used to develop prediction models.

As part of an Australian public hospital innovation's project, an ML model was developed to predict risk of readmissions within 28 days in emergency unit, using patients' demographic details and medical treatments/diagnosis from past admission. The project work was conducted to optimise an existing risk-prediction tool developed by the hospital. Case study 2 in section 8.3.2 discusses the model in detail, where the objective is to optimise the identification of risk readmission model used in the emergency unit.

8.3.1 Case Study 1: Identifying healthcare need and health services usage pattern using Machine Learning

8.3.1.1 Problem Identification

The insurance industry is facing a decline of membership numbers for private health products and services. To improve clients' engagement, an Australian health provider that provides health insurance products and services to Australian military families such as serving Australian Defence Force (ADF), ex-serving and the wider defence community conducted a research project. To meet this objective clients' health requirements were closely investigated.

Currently, serving (ADF) members and their families, and ex-serving members and families have different health requirements because of the effects on families and members of military services work and lifestyle. Often, families face mental problems due to isolation, relocation to regional or remote areas, and their lifestyle. Therefore, the objective is to identify key health diseases or conditions that are present in both the segments in order to better meet their health requirements. Furthermore, identification of health services use patterns for different segments can lead to better product development and support. Hence, service use patterns are predicted for each segment using service-usage model to assist with product development.

8.3.1.2 Method

With respect to the objectives of the organisation, an approach has been designed whereby research and data analytics are combined using machine learning. Approach was implemented in three stages, where initially literature review is done to understand lifestyle of clients. As the insurance company provides health services to Army professionals and their families, their mental and physical aspects were researched. Further, the internal database was analysed to determine the health pattern of different segments of clients and compared with the research findings. Finally, a prediction model was developed to predict the main health issues of each segment, utilizing their demographic and previous medical history details. The use cases can be integrated in existing products/services to improve new membership numbers or to upgrade the company's products/services. The three stages are discussed in detail below.

8.3.1.2.1 Literature review

A wide number of (50+) journals related to the health industry, clinical researches, defence white paper, reports from external sources such as AIHW, and previous research papers were considered for the literature review. All findings were collated to understand the lifestyle and major health problems of defence force members and their families. The aim of this research work was to know the lifestyles and challenges faced by AFD members and their families which could make them susceptible to specific health problems. Moreover, the literature review revealed the differences and similarities between the findings of previous research and those derived from the data analysis carried out in this study. Research helped in collecting predictors through hypothesis/research findings which were utilised during the data modelling stage. Demographic details such as location (e.g., remote or rural), marital status etc., are the key identifiers in this section as they an important role in defence families' wellbeing.

8.3.1.2.2 Data mining

As a next step, clients were segmented into four groups i.e. (I) ADF member (II) Spouse of ADF (III) Ex-serving members and (IV) Ex-serving spouse, based on their lifestyle and work profile. Company's internal database is then analysed to retrieve each segment clients' demographic details. A novel approach is introduced to solve the problem, where clients' past episode level hospital information is analysed and extracted to determine top health conditions for each segment. Episode level hospital data from the past three years has been used for each segment, where the primary and secondary diagnoses of every episode for each client were examined to determine the top ICD10 codes (diseases) for each segment. Thus, the main health conditions were identified from the data mining process, which showed few consistent similarities with the research findings of study and few gaps.

8.3.1.2.3 Data modelling using Machine learning

In order to predict the key health conditions of each segment, a predictive model is designed and developed using the machine learning approach. A prediction model is designed whereby the main health conditions of each segment are the 'target' and clients' demographic details and episode level information the 'features' used to predict main significant health conditions of each client. A service usage model is also built to assist the company to understand the pattern of health services used by each segment for specific health conditions. The development of the model is described and discussed below.

8.3.1.2.4 Development of a predictive model to identify health needs

Machine learning is a method of data analysis that automates analytical model building (SAS). A predictive machine learning approach is designed and proposed where the outcome of key health conditions is predicted '*y*' as yes or no using '*x*' features or variables. The development of a model involves the following steps.

- IDENTIFYING PREDICTOR 'Y': In order to identify the top health needs of different client segments, a predictive machine learning approach is designed and proposed where the binary outcomes of key health conditions are predicted as 0 or 1, "0" being unlikely (or no) and "1" being highly likely (or yes). To determine the main health issue of different segments, five health conditions are derived from step I and step II that include circulatory, respiratory, musculoskeletal, mental and, diabetes II conditions. These five diseases are target '*y*' in the model; also, predictive data models are created for each disease.
- FEATURE ENGINEERING AND ANALYSIS: Client information such as demographic details, lifestyle features and domain-specific insurance issues associated with the patients' health needs are analysed and extracted from the internal database. Additionally, clients' episode level hospital medical history has been taken out to closely relate their health requirement. Various significant features 'x' such as Gender, Segment, Location, Rank of ADF/Ex-serving, Age, Income, Remote location, and episode level-hospital information such Length of hospital Stay, Chronic Disease diagnosis details are selected as a result of extensive feature engineering and data analysis, to develop a predictive model.

8.3.1.2.5 Data collection and preparation:

Data from past three years was obtained from company's internal database using SQL server and Yellowfin BI tool. The data used for modelling purpose is cleansed, prepared, normalized and scaled using R and MS Excel to meet the prerequisites of the ML algorithms process. Further, the dataset (~150000 records) is partitioned into two sets: 80% for training and 20% for testing or validation.

8.3.1.2.6 Models for ML and validation

Multiple ML supervised algorithms were assessed and selected based on their nature (i.e., regression or classification, the type of problem, the size of the data and the domain-specific limitations. A number of algorithms are adopted for this problem, such as logistic regression. Random Forest, Gradient boost machines, SVM and Decision tree algorithms are used to build the models. A brief introduction to each algorithm is given below.

- Logistic regression: Logistic regression is a supervised learning method for fitting a regression curve, $y = f(x)$, when y is a categorical variable. The typical use of this model is to predict y given a set of predictors x. The predictors can be continuous, categorical or a combination of both. It works well for the diagnosis of medical problems.
- Decision Tree: A decision tree is a hierarchical model comprising discriminant functions, or decision rules, that are applied recursively to partition the feature space of a dataset into pure, single class subspaces [127]. A decision tree analysis is a divide-and-conquer approach to classification. Decision trees works well if one wants to discover features and extract patterns in large databases that are critical for discrimination and predictive modelling. Utilizing these characteristics, along with intuitive interpretation capability, are some of the reasons that decision trees have been used extensively for both exploratory data analysis and predictive modelling applications for more than two decades.
- Random Forest: The Random Forest is made up of several decision trees; each decision tree will be full grown and does not need pruning. The greater tree, the more accurate the result will be, and it will not over-fit. The random forest algorithm is applied for the overall estimate as it has the advantage of automatic feature selection [128].
- Gradient Boost Machines: In GBM, the learning procedure consecutively fits new models to provide a more accurate estimate of the response variable. This algorithm is used to ensure that the new base-learners are maximally correlated with the negative

gradient of the loss function, associated with the whole ensemble [129]. The loss functions applied can be arbitrary, but to give a better intuition, if the error function is the classic squared-error loss, the learning procedure will result in consecutive error-fitting. In general, the choice of the loss function is up to the researcher, with both a rich variety of loss functions derived so far and with the possibility of implementing one's own task-specific loss.

- Support Vector Machines (SVM): are systems which use hypothesis space of a linear functions in a high-dimensional feature space, trained with a learning algorithm from optimisation theory that implements a learning bias derived from statistical learning theory [130]. SVMs are a new promising non-linear, non-parametric classification technique, which have already shown good results in medical diagnostics, optical character recognition, electric load forecasting, and other fields [4].

All models are executed and compared for accuracy and speed. A number of model fitness tests, such as ROC, Accuracy test, Anova test and McFadden R^2 index, were also performed to select the best predictive model. The model was developed using logistic regression for the main five conditions are selected because it gives 80%+ accuracy and provides easy and useful interpretation of results for company executives.

8.3.1.2.7 Model 2: Predictive model for health usage pattern

To understand health usage patterns, a predictive health service model is designed and developed which identifies the pattern of health services used by each segment. A health usage model consists of health needs model outcome and predict health services pattern for all top needs of client segments. For example, health services pattern for ex-serving personnel will predict the usage of ancillary services, medical and hospital services for circulatory diseases, diabetes type II and musculoskeletal conditions. This model allows the insurance company to customize products and apply strategy to meet their clients' needs. The health services usage model shows 69%-79% accuracy for three main health conditions: musculoskeletal, circulatory and respiratory. Findings show that clients make more hospital visits with longer stays when diagnosed with musculoskeletal and respiratory diseases. Physiotherapy is widely used by both the segments if diagnosed with musculoskeletal/respiratory diseases. Table 8.1 shows the results for circulatory and musculoskeletal conditions from the prediction model.

CHRONIC CONDITIONS	Circulatroy		Muscukloskeletal	
SEGMENT	ADF/Ex-serving	Civillians	ADF/Ex-serving	Civillians
HEALTH SERVICES				
Hospital visits	22%	19%	33%	28%
Length of stay in hospital	17%	13%	35%	39%
ICU days	38%	40%	21%	22%
Psychcare days	31%	23%	75%	86%
Palliative days				
Alternate therapy			9%	
Ambulance				
Chiro/Osteo				
Chronic Disease Management		13%		
Dental		0%		
Devices				
Health and Wellbeing	0%			0%
Optical			10%	
Other	0%			
Pharmacy	0%			
Physiotherapy	12%	14%	69%	62%
Accuracy (Results)	**78.10%**	**79.50%**	**69%**	**69%**

Table 8.1 Results for circulatory and musculoskeletal conditions from the health use prediction model

8.3.1.3 Discussion

A novel model is built for the insurance sector, where the clients are current or former ADF personnel - active, reserve or retired. Lifestyle and demographic details such as gender, age, location, income and hospital admission-related medical history such as length of hospital stay, and chronic disease diagnosis are combined to predict the key health conditions of clients. The approach designed here combines research and data science to develop a novel model. Research is performed to comprehend the key health conditions of ADF and ex-serving members and families, and to investigate their different lifestyles and evaluate their impact. Findings from the literature review indicate the main health issues of members, spouses and children. Results show that musculoskeletal, mental and behavioural issues are the main health problems for ADF personnel, both current and ex-serving.

As a next step, extensive data analysis and discussions with executives and interdisciplinary authorities and insurance domain experts were undertaken to understand the health database, and the key parameters of their current products and strategy. Finally, useful health-related information was extracted from the company's internal database. Feature engineering is also performed to extract effective variables from the client database to help predict the output. Five supervised machine learning algorithms are applied in parallel; of these, the logistic regression and gradient boost approaches performed best in terms of accuracy. The gradient boost approach yields the best results for health outcome predictions and this why it was chosen to build model. Logistic regression was chosen as it achieves 80%+ accuracy and offers clear interpretation of results. Both algorithms were used to build prediction models for the company.

The execution of the model revealed various interesting findings, particularly for military personnel when compared to civilians. Model implementation suggested different use cases to predict health conditions. Similarly, an advanced health service model is built in order to understand the pattern of health service used by different client segments. For instance, the health service usage pattern indicated that ADF personnel (current and ex-serving) used health services far more than did the civilian clients.

A large dataset was used for the development of the model, combining knowledge and comments from multidisciplinary teams within the organization. Because hospital details are included in the model, this is an innovative step in the insurance industry. The models can assist the company to analyse its current products and customize them according to the use cases produced by the model. Additionally, the model can be developed as a tool to predict the health needs of new clients or can be used as a use case which can help the company to revise or develop new products to meet each segment's needs. Hence, the aspect of ML which indicates that "data models do not provide correct interpretations" is solved here by providing use cases and tool development.

8.3.1.4 Limitation and Future Scope

Data models are developed using clients' primary and secondary diagnoses obtained from health data, along with their demographic and work information from internal data to predict the health conditions. Only chronic conditions are considered from episode diagnosis. Few medical issues can impact others or reveal new health conditions. Thus, actual knowledge about medical conditions and their relationship can lead to realistic models with better impact. Health product development can also be done using data modelling.

8.3.2 Case Study 2: Optimizing the model for Identifying risk readmission using ML

8.3.2.1 Problem identification

Some hospital readmissions are unavoidable, but some can be reduced by identifying risk of readmission, which can subsequently save money, hospital beds and time. Alternatively, low-risk patients can receive appropriate care via community services, GPs and other health care services [60].

Pertinent to the issue, an Australian public hospital initiated an innovation project and developed a risk -prediction tool to identify the risk of unplanned readmissions to the emergency unit. This model was developed as a prediction tool using machine learning approach and clinician's knowledge and expertise to determine the main indicators of unplanned readmission to the emergency unit. In the next phase, the validation and optimisation of the prediction tool is conducted to achieve following

- Validation of the existing model by verifying the correct dataset
- Improving the model's accuracy by using a large dataset
- Building an iteration model to compare or predict new readmissions using ML.

8.3.2.2 Solution Outline

To optimise the ML model, two major methods are investigated. First, a large amount of data is used to re-execute the existing model; second, it is validated by partitioning the dataset into 70%-30% for training and testing the model to improve its prediction accuracy.

Another way to analyse an internal database is to find more affective factors for patients' readmission, and build a new model using strongly associated factors identified by the machine learning approach. This model makes use of statistically significant variables to predict the risk of a patient being readmitted. Thus, adds more reliable and visible insights on readmissions. After performing validation and optimisation, the best model can be used as a prediction tool or findings can be used as a basis for another optimised model.

This section discusses the complete approach used to validate and optimise the existing model and suggest steps for developing a new model.

8.3.2.3 Detailed approach to solve the problem

Validation and optimisation approaches are used to improve accuracy. This section discusses each process in detail; this is followed by the development of the model.

8.3.2.3.1 Validation of current model

As part of the validation process, the model is internally verified using clinical tools. The output from the verification shows that the dataset chosen to create the model has few errors. The errors are corrected using the clinician's knowledge and verified again. Finally, the correct data is extracted from the database and prepared for modelling.

8.3.2.3.2 Optimisation

Optimisation approach is designed so that the database is again analysed to identify statistically-strong factors that contribute patients' readmissions. The new factors are used to develop a model and are validated using a large dataset for accuracy. The optimisation process involves the following steps:

- Readmission analysis by performing feature engineering and data mining: In order to understand and determine the salient factors contributing to unplanned readmissions to a hospital, extensive data mining is performed. New factors include new variables such as admitted speciality, discharged speciality, private insured patient, total number of diagnoses in previous episode, principal diagnosis etc.
- *Machine learning limitations:* Furthermore, a literature review of hospital readmissions is conducted and the characteristics of machine learning models are explored to resolve the issues in the existing model. There exist a number of shortcomings in the application of ML in the health industry. Issues such as errors in the dataset and small dataset used for modelling are identified in the existing model.
- *Development of a new prediction model for readmission risk.* A new model is developed to evaluate most of the unplanned readmissions occurring within 28 days, taking into account the patients' demographic details, medical complexity and diagnosis/treatments received during their previous hospital admission. A predictive machine learning approach is designed and proposed where readmission risk 'y' is predicted as 'yes' or 'no' using 'x' features.

8.3.2.4 Model Development

8.3.2.4.1 Data preparation

The data extracted consisted of one year of readmissions to a specific hospital within 28 days. Also, patient records with no readmission and readmission >28days are considered for the model. The hospital admission details for each patient were collected, and descriptive statistics, correlation and covariance were determined to select the statistically significant variables. Finally, the dataset was cleansed, normalised and scaled.

8.3.2.4.2 ML modelling and validation:

A new model is developed using the patient's demographic details such as age, gender, rural/remote location, ethnicity, private insurance etc. However to optimise the model, the maximum amount of available information is required. Patients with complex conditions are assessed using a formula that evaluates their diagnosis (ICD10) and mark their complexity as "complex" or "not complex" (approach suggested by hospital's pharmacist). Additionally, patients' previous admission details such as total number of diagnoses, treatments, admission specialty, discharge specialty, principal diagnosis and hospital stay details are also taken as input for the optimised model.

8.3.2.4.3 Logistic regression algorithm is used to build predictive model

The prepared dataset (~10 records) was partitioned into 70% training and 30% testing data, respectively. Logistic regression is used to build the model. R was used to write the logistic regression code and the prepared training dataset was used to run the model. Testing was performed, and results showed a 75% accuracy. Multiple model fitting tests were performed to select the best model. The final model showed75% accuracy and 65% ROC.

8.3.2.5 Discussion

The existing model is a part of an innovation project, where a prediction tool is developed to identify risk of readmission within 28 days of patients being discharged. The existing tool is a result of combining machine learning approach and pharmacist' or clinicians' knowledge and recommendations. The model shows good accuracy. In the next step of the project, the aim is to improve the model accuracy by using validation and optimisation techniques.

The approach involves a design where model limitations are identified first and clinical validation is performed to determine what needs to be improved.

An optimisation approach is taken to build the solution model and maximise its accuracy. A new model is developed in order to build a machine learning model using statistically strong variables to obtain new insights on risk readmission. The model achieves 75% accuracy for prediction and 65% for ROC, which is considered good in the medical the domain. New model outcomes revealed strong variables which can either be included in the existing model to enhance the model accuracy and interpretation or be utilised to build a revised tool. The existing model was limited in that it used a small dataset. Also, the sample dataset is verified using a clinical validation process to determine whether the dataset contains the correct set of patients. Any identified errors are fixed in order to obtain correct data for the model.

The work done in the project demonstrated the optimisation of a ML model. It also discusses and explains the various machine learning limitations and biases. A good accuracy model can reveal patterns which may or may not be useful unless they are clearly interpreted or optimised. The optimisation of a model based on new insights can be seen as a step toward a more effective model. New insights can be added to the existing model to either create a new model or just to improve the overall accuracy. It is essential to validate the model internally, especially in the health space, to understand the underlying data and patient story. A data model depicts pattern from the data so it essential to have the right set of data so that the outcome reflects the reality.

8.3.2.6 Limitation and future work

The model development is a result of a short period of work undertaken for the project. Although the current model has used ~10k records, ML is better when trained on a very large dataset. Hence, 4-5 years of data can be utilized to improve the accuracy. Clinical validation is proposed as a future work. Due to data ownership and access issues, only a limited number of data sources have been used to create the model. As a future project, the model could be improved by including information from multiple hospital data sources.

8.4 Roadmap for health prediction models

As a result of work done in health industries projects, a roadmap has been developed based on the identification of health project requirements and the limitations of the health industry such as the lack of data expertise and knowledge gap between data science and health science. Figure 8.1 presents a roadmap, using machine learning, for the development of a prediction tool for medical problems or medical-related projects.

8.4.1 Roadmap

For health-related projects, it is essential to conduct a literature review that will give a comprehensive understanding of the client or patient. It is also beneficial to obtain clinicians' knowledge regarding medical issues or procedures. To build a prediction model, the normal process is to take steps to create a roadmap. These are:

8.4.1.1 Identifying target:

For medical problems, predicting a target "y" is clearly outlined but for some projects the target and its type needs to be identified. A target can be continuous, ordered value or categorical. Predicting target defines the machine learning algorithms that can be utilised to build model. For example, for continuous a target, MLAs such as regression, clustering, time series can be used. Whereas, for a categorical target, decision tree, random forest, GBM, or logistic regression can be used. Multi-variate is normally utilised for ordered values.

8.4.1.2 Feature engineering and data analysis:

In order to understand the important features contributing to the problem, i.e., “y”, feature engineering is used. Feature engineering is a term used for the combination of domain knowledge and business rules to analyse data and identify influencing parameters which may impact the target variable. To identify “x” or features, extensive data mining is performed to determine the salient features responsible for predicting the target.

8.4.1.3 Data collection and preparation:

Data collection and preparation is an integral part of data modelling. Data is extracted using business rules or medical data flow. Generally, a large amount of data is desirable for model development, although in the health domain, an appropriate set of patients or clients is more important. Thus, data is collected and prepared for data modelling, where missing values are replaced with mean or median values. Data is normalised and scaled to obtain values at the same level; i.e., high numerical values can be converted to range or categorical to reduce them.

8.4.1.4 Model development:

Depending on the target variable, a number of supervised algorithms are chosen. For the health industry, logistic regression, GBM, SVM has proven good results. However, a number of algorithms can run in parallel to check the model fitness. Sample dataset is divided into 70%-30% to form training and testing datasets. The training dataset is used to train the data, and 30 % is used to predict the model accuracy.

8.4.1.5 Model validation:

There are a number of model fitness tests such as ROC, Accuracy test, Anova test, and McFadden’s R^2 test.

In the health domain, after obtaining accuracy results, clinical validation is a next step. It is important to validate the dataset and determine whether or not the patient data used for modelling is correct. With prediction outcomes, it is easier to clinically validate the data. If the sample dataset is correct, a large amount of data can be used to re-execute the model to obtain better accuracy. However, if the sample dataset is not correct, the model can be optimised by performing feature engineering using appropriate business rules or medical data flow. The whole process can be repeated to obtain good model accuracy.

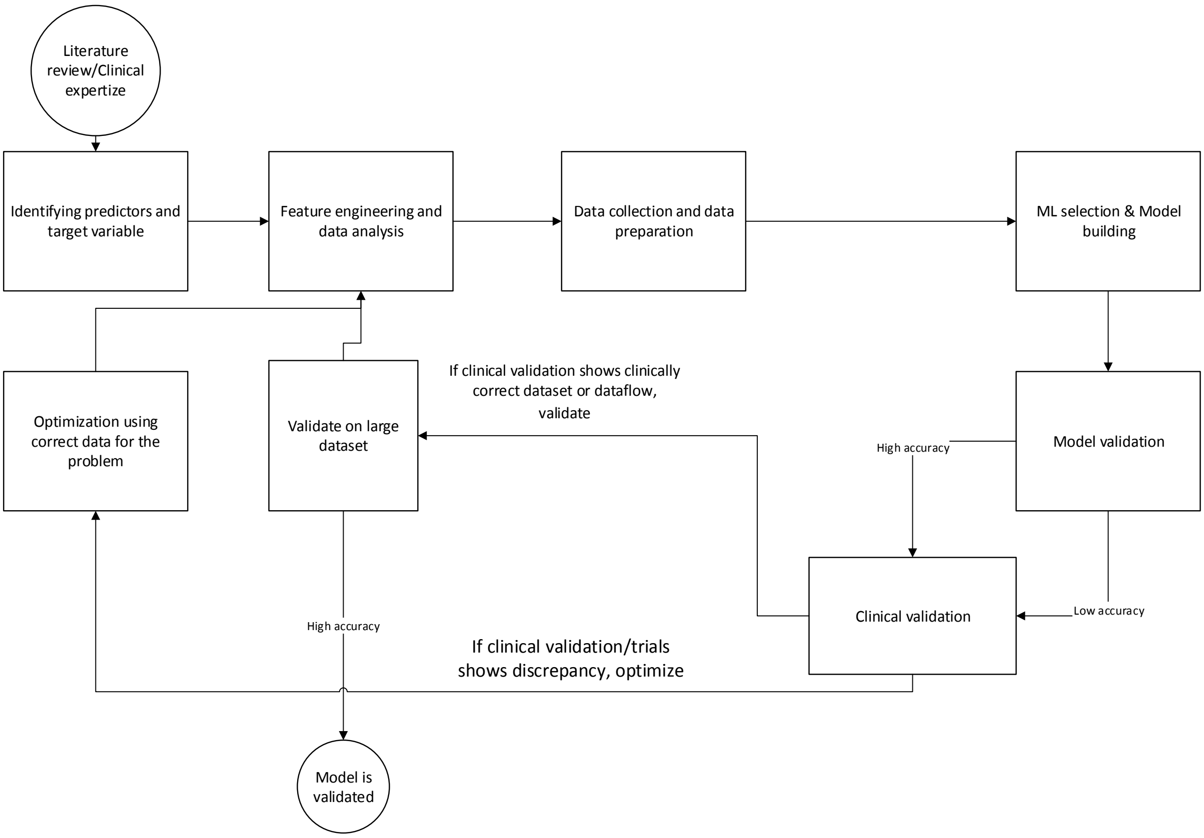

Figure 8.1 Roadmap for prediction model development in health industry

8.5 Conclusion

Big health care data offers tremendous potential for providing information that can be acted upon quickly. Advanced technology applied to big health data can support and improve timely decision-making. However, big health data is useless unless it is analysed or interpreted correctly to identify useful patterns or knowledge. Hence, machine learning and AI applications are gaining much attention in the health care domain. A number of advanced tools and systems are currently facilitating the medical process and patient care. Having said that, there are questions regarding the clarity and outcomes on ML models. This study discusses ML implementation in detail and several major aspects of ML such as model biases, methodological limitations, interpretation of the models, with a viewpoint to answer ML limitations. Also, a roadmap is developed in this study which is based on an understanding of the health industry's requirements in terms of data science, and the current shortcomings of the health industry such as lack of data expertise and knowledge gap among health sector and data professionals.

The work performed for the organizations shows that clinical validation is key to understanding the model and the story it yields. For an ML model, data is a key requirement. The more accurate the dataset, the better will be the outcomes. Thus, in the health industry, a problem can be better understood by optimizing the models and validating them by using internal clinical tools/software and meeting the data requirements. Models can be improved by optimisation activities such as applying them to large data or creating new models for the same problem and adding new finding to the model and iterate it to improve the accuracy.

Model interpretation can be done by suggesting ways to use the model insights or implement it as a tool. Thus, ML models are not just about building models with high accuracy, however correct interpretation of a model and usage must be answered.

Chapter 9

9 Future work and Conclusion

This chapter provides:

- The issues and problems of local demand food, involvement of agro prosumers and sustainability are presented in form of agro-prosumer community groups.
- Solutions proposed in this dissertation to address the problems.
- The limitation of affective design evaluation using fuzzy regression is addressed using a social media data analytical framework and ML approaches.
- Conclusion followed by future work.

9.1 Introduction

In this literature, as a key research project, the concept of a community agriculture group has been used to represent alternative food network. The CSAs are discussed widely and multiple approaches have been proposed which addresses different challenges associated to it, even the concept of carbon and produce sharing community network group have not yet been developed with a detailed structure. Also, the prosumers involvement in community agriculture group with an aim to create carbon/produce exchange platform have not been comprehensively investigated; only very few research studies focus on the involvement of the local prosumers in the community agriculture groups or urban prosumer groups in the carbon or produce sharing network.

The prosumers involvement as community groups has been acknowledged, and interest in it is increasing due to local food demand and quality of food. How to create a prosumer group? The definition of agro-prosumers, the recruitment of new members to the existing groups, managing members and their goals and providing a platform to trade the produce while maintaining transparency and traceability, are important factors to consider. In past literature, approach to comprehensively address the associated problem have not been proposed and developed.

Hence, to improve the participation of prosumers and overcome this barrier, this study identified are four key milestone in developing an Agro-Prosumer Community Group (APCGs) to form, manage agro-prosumers group and address their needs.

Another interesting work has been carried out in the area of affective design as part of this PhD journey. In evaluating affective design perception, fuzzy regression techniques have been used widely to address issues such as the accurate measurement of uncertainty in human perception. However, until now, none of them has utilised social media data to collect design perception and evaluate it using Fuzzy techniques or ML analytics. This study proposed a new social media analytical framework (SMAF) to collect and analyse social data for the purpose of product design.

In the following section, recapitulated the issues that are identified and subsequently addressed as part of this thesis. Section 9.3, highlights the thesis contributions to the literature research by addressing these issues. Proposed future work is discussed in section 9.4, and section 9.5 concludes the chapter.

9.2 Problems Addressed by This Thesis

This thesis targeted two problem areas and one applied research knowledge in healthcare space. Firstly, the concept of community agriculture or a prosumer group has been discussed with approaches such as urban prosumers, community gardens, CSAs. Although there are approaches currently proposed by different studies which investigate community group's concepts, however none develops a comprehensive framework to resolve the research problems addressed in this thesis. A new concept of developing agro-prosumer community groups APCGs and solving four key problems associated with community groups have been addressed in this dissertation. Secondly, a new framework utilizing social big data and analytics is used to address uncertainty evaluation in users' perception. A case study is used to discuss the solution. Finally, machine learning approaches are utilized in health industry to build prediction models using big health data and a discussion is provided that addresses various related issues. As mentioned in Chapters 1 and 2, the research in the area of community agriculture shows a gap in prosumer community group concept particularly to share produce and carbon in a trustable network. Hence to provide a community network platform for prosumers following issues have been identified:

- Problems with defining and clustering prosumers into APCGs
- Issues with new prosumers recruitment to APCGs
- Complications in negotiating and managing different goals within the APCGs

- Difficulty in assessing carbon content and lack of a trustful platform for sharing/trading carbon/produce with the network and buyers.
- Issues with evaluating uncertainty of affective design perceptions.
- ML model limitation aspects in health big data are discussed

As discussed in Chapters 1, 2 and 7, it is evident that research studies on community groups are still susceptible to these issues; thus, the drivers of this research are these four issues which form basis of subsequent development of an innovative framework.

9.3 Contributions of this study

The key contribution of this dissertation to existing literature is that it addresses four main issues relevant to the development of an agro-prosumer community-based approach which manage the prosumers in local food- and carbon-sharing networks. In order to solve the aforesaid research problems outlined in Chapter 2, an APCG concept is developed via presenting four building sub-frameworks, which form the core contribution of this dissertation to existing literature. In addition, a new social media analytical framework was developed to collect and analyse uncertainty in affective design perception which is discussed in detail in chapter 7.

Thus, the principal contributions of this thesis include

- APCG concept framework
- APCG recruitment framework for new Agro Prosumers
- APCG goal management framework
- Carbon assessment and carbon-token framework
- SMAF framework to evaluate the affective design uncertainty
- Discussion of ML prediction model in the health domain using a roadmap

9.3.1 APCG concept framework

The term 'Agro-Prosumer Community Group' for sharing/trading carbon and produce is coined in this dissertation. The current efforts of community groups and prosumer involvement has been investigated in the area of community agriculture. Food prosumers, sustainability and the approaches used for prosumption, CSAs, carbon farming and blockchain platform for trade in community groups, have been discussed in the literature. However, the studies on prosumer community groups did not proposed approach for characterizing and defining APCGs for the sharing of produce and carbon. Moreover, the comparable approaches applied to existing community group projects have attempted to involve farmers and the community in the form of farmer-member relationship; thus, they call for farmers' problems but do not involve prosumers or gardeners in community groups, and do not give much attention to managing prosumers for mutual goals and trading purposes. However, this research suggests new aspects of APCG characterization and definition, which are not covered in past studies. This thesis investigated all such research studies to obtain additional understanding.

In this thesis, defined the APCG concept, analysing the prosumers produce-sharing behaviours (seasonal produce pattern) and clusters the prosumers with similar production-sharing behaviours and finally operationalise the APCGs. The key highlights include

1. Prosumers' dataset is considered; this includes production data, which shows yearly or seasonal production profiles for average winter/summer season or year-long production.

2. Prosumer data set outliers, i.e., Clustering algorithms is used to identify the prosumer yield outlier by identifying prosumer's that are further away from the standard yield behaviour than the deemed reasonable. To cluster prosumers, hierarchical clustering method is used.

3. Hierarchical clustering grouped the profiles based on the homogeneous production amount.

4. Finally the clusters are optimised to obtain APCGs by splitting the large clusters or merging the neighbouring small clusters, thus ensuring that each APCG supplies a given APCG threshold produce, and have at least minimum number of prosumer profiles in each APCG.

5. APCG pre-condition criteria are identified by analysing the attributes of the APCGs. These pre-condition criteria can serve as the guidelines for prosumers future recruitment.

6. Using R environment, proposed framework was analysed with 300 prosumer produce records, initially identified 8 prosumer clusters which were optimised into 6 APCGs taking into regard that each APCG supplied their minimum threshold production quantity and each APCG consist of at least 10 prosumers. To best of our findings, this approach is not proposed in the current literature which defines and characterizes the APCGs; hence, this work makes a novel contribution.

9.3.2 New prosumer engagement framework

As mentioned in Chapter 2, the existing literature does not contain studies that focus on the structure of APCGs and their function as sharers of produce and carbon; the engagement of new prosumers to the APCG can be seen as a next step to providing a seamless structure for APCGs. Additionally, current community agriculture groups also lack comprehensive approaches for recruiting new prosumers. For example, little consideration is given to the management of prosumers for the purpose of recruitment by, for example, evaluating prosumers before recruiting and later analysing to categorise the complying and subsequently re categorizing the non-complying prosumers. Thus, the approach presented in this study can be used in CSA groups with minor revisions.

Chapter 4 provides a framework for prosumer recruitment in which the new prosumers' real commitment, and over the period real production are evaluated prior to the prosumers' allocation to appropriate APCGs. The key highlights include:

1. The notion of having an "evaluation period," during this period prosumer produce performance is observed with objective to assign the final APCG, is discussed. This framework, suggested an ongoing prosumer produce evaluation over two consecutive seasons for a seasonal crop or two years for annual variety. The prosumer is analysed in each season slot and at the end of their evaluation window.

2. The variety and amount of produce that a prosumer contributes is evaluated against the pre-requisites of the chosen APCG. The presented approach identifies the prosumers success and reliability against their chosen APCG, and the prosumer's chances of being recruited to an APCG is computed.

3. MATLAB prototype environment which analysed the 50 new prosumer data to validate the proposed prosumer recruitment framework. To the best we know there are no current literature which has proposed a framework for the recruitment of agro-prosumers to the APCGs, making our thesis novel in this field.

9.3.3 Community based Goal management framework

As discussed in literature review of APCG Chapter 2, there are no existing research on APCGs and on community agriculture which provide an approach for managing the different goals of a community-based, shared network produce; APCGs.

These problems are dealt with in Chapter 5 where a framework for goal management is provided that effectively reaches a compromise between conflicting goals of APCGs. The key highlights of this framework are:

1. The diverse goals of APCGs are recognised and later classified as definite and flexible constraints. The constraints prioritization is taking into account their relative importance which is identified by adapting the MCGP goal- ranking technique.

2. Goals are prioritised and conflicting goals are negotiated to find an optimised goal solution, requiring varying amounts of resources and alterations to parameters in order to achieve all the goals.

3. Verification of this proposed goal-programming solution is conducted by defining it as a linear goal programming issue, which is solved using LINGO software. To best of our knowledge there is no comprehensive work in the current literature on the APCGs' goal management for community agriculture group, this proposed solution framework is novel in this research area.

9.3.4 Agro-prosumers carbon capture evaluation and blockchain framework

As discussed in literature review of APCG Chapter 2, the main reason for establishing an APCG is to contribute to environmental sustainability and meet local food & safety demand. To date, previous studies have not discussed prosumers carbon assessment and the building of a trustworthy trading platform for community group. In addition, the literature on other issues such as carbon farming, food security, and agriculture has used blockchain technology. However, APCGs combine three tasks in one: carbon farming, trading and production sharing network. These, hence the difference between the APCGs and other food networks, making the former unique.

This thesis develops a framework in order to address the issues identified in the literature, for assessing prosumers' carbon capture and building a platform for APCG trade as discussed in Chapter 6. Here, the carbon capture from trees are only considered to ease the carbon calculations as part of prosumers carbon assessment.

The key features of the framework are:

1. The initial phase carbon absorption is computed using prosumers' tree information. A general carbon calculator method is used to calculate carbon absorption by using height and diameter of a tree. For each prosumer in an APCG, the carbon absorption is computed and combined to obtain the total quantity of carbon absorption. It is suggested that usually, one tree absorbs on average about 50 pounds of carbon.

2. Next, the DEFRA method is used to calculate carbon emissions by prosumers. To make prosumer groups sustainable, prosumers' carbon footprint is evaluated and deduced from their total carbon absorption. Surplus carbon captured is converted into carbon credits for trade.

3. The carbon credit is converted into a carbon token electronically which can be traded by APCG and organization or government bodies.

4. The proposed carbon token framework pseudocode is developed using blockchain and a structure of Database is discussed to implement APCG.

The current literature pertaining to APCGs contains no studies analysing this issue, making this thesis a novel contribution.

9.3.5 SMAF framework for evaluating affective design perception using Fuzzy regression techniques

The final major contribution of this thesis is the development of a unique framework which extracts data from social media platforms such as Twitter and Instagram. The solution proposed is a social media analytics framework (SMAF) to predict uncertainty in affective design, integrating fuzzy regression. Until now, users' affective design perceptions were either collected through surveys or predicted by models. In both the cases, a wide range of users were not able to participate and, moreover, all design features were not used for surveys or prediction models due to surveys being lengthy, attracting less interest from the users, or having technological limitations. Thus, the extraction of a massive amount of dynamic data for affective design evaluation is a unique contribution of this study.

The social big data texts related to users' design opinion and perceptions are retrieved and transformed into structured data. This design data is analysed and evaluated to accurately measure the uncertainty present in users' opinions.

In order to demonstrate the effectiveness of the framework, a case study of luxury family is conducted and the results are reported.

Chapter 7, we discussed in detail the framework development and the case study. The proposed framework has the following highlights

1. SMAF development: The SMAF is used to extract and analyse affective design data from social media platforms such as Twitter, Facebook and Instagram. The framework is developed using PHP and SAS tools, which extract keyword-based tweets from Twitter. Then it transforms the unstructured data into structured data. The structured design data is then analysed using an automated program developed using NLTK for a five-point scale evaluation.

2. Fuzzy regression techniques: Various fuzzy regression techniques were studied to evaluate uncertainty in design texts.

3. Evaluation using a case study: Affective quality perceptions are evaluated using the fuzzy regression techniques, and demonstrated using a case study of luxury family car such as BMW, Mercedes, Audi, Lexus, and LandRover. A case study of affective design of automobile was used to demonstrate the application of SMAF. The colour feature of a car is used as the design attribute to predict the affective quality of car models.

This framework makes a unique contribution to the area of affective design evaluation. To the best of our knowledge, fuzzy regression has, to date, not been used to explore social media data related to affective design.

9.3.6 Machine learning prediction modelling aspects in health space

Big data and ML techniques are evolving in the health industry and are now used to develop tools and automated systems to predict diagnosis, treatments, risks and services for better support and quick decision-making. Along the same lines, ML is used to meet the objectives of a private health insurance provider for their research project. Various studies have addressed the use and limitations of ML in the health industry in order to improve the efficacy and technical aspects of machine learning. This study gives an Outline of prediction models developed during the internship and short placement period, with a view to understanding the features and limitations of ML.

The highlights of this study are

Case study 1: A novel model is developed to predict patients' main health conditions based on their lifestyles and historic medical data. To best of my knowledge, this is a unique model created for an insurance company where data on a client's lifestyle and medical history are combined to predict their health needs and services pattern.

Case study 2: A ML model is developed to predict risk of readmissions to the emergency unit within 28 days. It makes use of patients' demographic details and medical treatments/diagnosis from past admission(s).

The contribution of this study is that it ascertains the limitations of current models and offers several proposals to address them. However, major revisions are required to address different problems as the health industry is unpredictable and complicated.

9.4 Future work potential

The research in this thesis is discussed in conferences and published in peer-reviewed international journals. Over the tenure of this research, three research papers are published in two peer-reviewed conferences along with three peer-reviewed international journals. Although the research has formed the core foundation of framework, we still can see future scope as discussed below

9.4.1 In this research, four key issues of developing APCGs are evaluated and corresponding solutions are proposed. However, part of the solution developed in this work can serve as an initial building block which can address the aforementioned issues. An example, the carbon capture assessment can form as a stepping-stone for the development of a fair carbon absorption method.

9.4.2 In this research, the definition of APCGs and the prosumer recruitment is based on the homogeneous production-sharing capability of prosumers and their geographical location; however, other categorical values such as the variety of prosumers' vegetation and tactical plans that affect the production are not in scope.

9.4.3 The goal programming model in a goal management framework can be extended to include mutual goals distribution among the groups.

9.4.4 The fuzzy regression method used in this study has the limitation that only one design attribute is considered. Whereas, the data collected from social media has many design attributes such as: "type" (sedan, hatchback, SUV), "rooftop" (sunroof, full sunroof, wide sunroof, no sunroof), "shape" (sporty, muscular), "kidney grille" (large, medium, small) and "overall outlook" (supercar, dream car, fantastic, quietest). In the future, we will develop a model to utilize multi-variables to evaluate uncertainty in affective quality. We will compare the multi-variable models with other methods.

9.4.5 The SMAF framework has been developed for Twitter data extraction and analysis only, and Instagram data is extracted manually. However, multiple platforms can be combined and data fusion can be created by researching data extraction from Facebook and Instagram and automating it.

9.4.6 The ML models in the health domain are discussed in order to understand the limitations and complications related to the health industry. The viewpoint can be used as a stepping stone for future model building tasks.

9.5 Conclusion

This chapter provides a summary of the work undertaken and documented in this dissertation. Firstly, we refer to the literature review that identified the gaps in extant research that prompted this research. We then discuss the contributions made by this thesis, and follow this with a brief description of the future work that can be carried out to extend the research work.

References

1. Antonucci, F., et al., A Review on blockchain applications in the agri-food sector. Journal of the Science of Food and Agriculture, 2019.
2. Cowie, A.L., et al., Impact of carbon farming practices on soil carbon in northern New South Wales. Soil Research, 2014. 51(8): p. 707-718.
3. Ahluwalia, P. and T. Miller, The prosumer. Social Identities, 2014. 20(4-5): p. 259-261.
4. Alban, O. and M. Étienne, Diversity and Struggles in Critical Performativity. The Case of French Community-Supported Agriculture. Management (Paris, France : 1998), 2019. 22(4): p. 537-558.
5. Jessen-Hannula, J., The urban Prosumer in Finland–a path to Sustainability. 2019.
6. Bechtsis, D., et al., A Blockchain Framework for Containerized Food Supply Chains, in Computer Aided Chemical Engineering. 2019, Elsevier. p. 1369-1374.
7. Dumbrell, N.P., M.E. Kragt, and F.L. Gibson, What carbon farming activities are farmers likely to adopt? A best–worst scaling survey. Land Use Policy, 2016. 54: p. 29-37.
8. Rosano, M. and D. Duncan, The Australian carbon farming initiative: early developments ongoing challenges. 2012.
9. Behnke, K. and M. Janssen, Boundary conditions for traceability in food supply chains using blockchain technology. International Journal of Information Management, 2019.
10. Blattel-Mink, B., et al., Beyond the market--New practices of supply in times of crisis: The example community-supported agriculture. International Journal of Consumer Studies, 2017. 41(4): p. 415-421.
11. Kosnik, E., Production for consumption: Prosumer, citizen-consumer, and ethical consumption in a postgrowth context. Economic anthropology, 2018. 5(1): p. 123-134.
12. Freibauer, A., et al., Carbon sequestration in the agricultural soils of Europe. Geoderma, 2004. 122(1): p. 1-23.
13. Caro, M.P., et al. Blockchain-based traceability in Agri-Food supply chain management: A practical implementation. in 2018 IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany). 2018. IEEE.
14. Rathnayaka, A.J.D., et al., Formation of virtual community groups to manage prosumers in smart grids. 2015.
15. Cone, C. and A. Myhre, Community-Supported Agriculture: A Sustainable Alternative to Industrial Agriculture? Human Organization, 2000. 59(2): p. 187.
16. Creydt, M. and M. Fischer, Blockchain and more-Algorithm driven food traceability. Food Control, 2019.
17. Mishra, S. and S.P. Singh, Carbon management framework for sustainable manufacturing using life cycle assessment, IoT and carbon sequestration. Benchmarking: An International Journal, 2019: p. 01-2019-0044.
18. Kamilaris, A., A. Fonts, and F.X. Prenafeta-Boldύ, The rise of blockchain technology in agriculture and food supply chains. Trends in Food Science & Technology, 2019. 91: p. 640-652.

19. Dabire, D.E.R., et al., OPERATIONALIZING AN INNOVATION PLATFORM APPROACH FOR COMMUNITY-BASED PARTICIPATORY RESEARCH ON CONSERVATION AGRICULTURE IN BURKINA FASO. 2017. 53(3): p. 460-479.

20. Seran, S. and M. Izvercian, Prosumer engagement in innovation strategies: The Prosumer Creativity and Focus Model. Management decision, 2014. 52(10): p. 1968-1980.

21. Nazarov, A., V. Shvedov, and V. Sulimin. Blockchain technology and smart contracts in the agro-industrial complex of Russia. in IOP Conference Series: Earth and Environmental Science. 2019. IOP Publishing.

22. Velasco, E., et al., Does urban vegetation enhance carbon sequestration? Landscape and Urban Planning, 2016. 148: p. 99-107.

23. Dong, H., B. Campbell, and A.N. Rabinowitz, Factors impacting producer marketing through community supported agriculture. PloS one, 2019. 14(7): p. e0219498-e0219498.

24. Howson, P., et al., Cryptocarbon: The promises and pitfalls of forest protection on a blockchain. Geoforum, 2019. 100: p. 1-9.

25. Veen, E.J., H. Dagevos, and J.E. Jansma, Pragmatic Prosumption: Searching for Food Prosumers in the Netherlands. Sociologia ruralis, 2020.

26. Hartmann, S. and S. Thomas, Applying Blockchain to the Australian Carbon Market. Economic Papers: A journal of applied economics and policy, 2019.

27. Tristram, O.W. and M. Gregg, Net Carbon Flux from Agriculture: Carbon Emissions, Carbon Sequestration, Crop Yield, and Land-Use Change. Biogeochemistry, 2003. 63(1): p. 73-83.

28. Rathnayaka, D., et al., Identifying prosumer's energy sharing behaviours for forming optimal prosumer-communities. 2011.

29. Tripoli, M. and J. Schmidhuber, Emerging Opportunities for the Application of Blockchain in the Agri-food Industry. FAO and ICTSD: Rome and Geneva. Licence: CC BY-NC-SA, 2018. 3.

30. Hanson, K.L., et al., Adults and Children in Low-Income Households that Participate in Cost-Offset Community Supported Agriculture Have High Fruit and Vegetable Consumption. Nutrients, 2017. 9(7).

31. Post, W.M. and K.C. Kwon, Soil carbon sequestration and land-use change: processes and potential. Global change biology, 2000. 6(3): p. 317-327.

32. Kim, H. and M. Laskowski, Agriculture on the Blockchain: Sustainable Solutions for Food, Farmers, and Financing. 2017.

33. Lin, J., et al. Blockchain and iot based food traceability for smart agriculture. in Proceedings of the 3rd International Conference on Crowd Science and Engineering. 2018. ACM.

34. Kolodinsky, J.M. and L.L. Pelch, Factors Influencing the Decision to Join a Community Supported Agriculture (CSA) Farm. Journal of Sustainable Agriculture, 1997. 10(2-3): p. 129-141.

35. Krul, K. and P. Ho, Alternative Approaches to Food: Community Supported Agriculture in Urban China. Sustainability, 2017. 9(5): p. 844.

36. Opitz, I., et al., Consumer-producer interactions in community-supported agriculture and their relevance for economic stability of the farm–An empirical study using an Analytic Hierarchy Process. Journal of rural studies, 2019. 68: p. 22-32.

37. Paul, M., Community-supported agriculture in the United States: Social, ecological, and economic benefits to farming. Journal of Agrarian Change, 2019. 19(1): p. 162-180.

38. Njundu, S., L.J. Moffitt, and A.L. Daniel, Stochastic Efficiency Analysis of Community-Supported Agriculture Core Management Options. Journal of agricultural and resource economics, 2001. 26(2): p. 417-430.

39. Samoggia, A., C. Perazzolo, and P. Kocsis, Community Supported Agriculture Farmers' Perceptions of Management Benefits and Drawbacks. Sustainability, 2019. 11(12).

40. Thomaier, S., et al., Farming in and on urban buildings: Present practice and specific novelties of Zero-Acreage Farming (ZFarming). Renew. Agric. Food Syst, 2015. 30(1): p. 43-54.

41. Kingsley, J.Y., M. Townsend, and C. Henderson-Wilson, Cultivating health and wellbeing: members' perceptions of the health benefits of a Port Melbourne community garden. Leisure studies, 2009. 28(2): p. 207-219.

42. Zeng, D., et al., Social Media Analytics and Intelligence. IEEE intelligent systems, 2010. 25(6): p. 13-16.

43. Rathore, A.K., P.V. Ilavarasan, and Y.K. Dwivedi, Social media content and product co-creation: an emerging paradigm. Journal of enterprise information management, 2016. 29(1): p. 7-18.

44. Yan, Z., et al., Big data fusion in Internet of Things. Information fusion, 2018. 40: p. 32-33.

45. Bello-Orgaz, G., J.J. Jung, and D. Camacho, Social big data: Recent achievements and new challenges. Inf Fusion, 2016. 28: p. 45-59.

46. Stieglitz, S., et al., Social media analytics – Challenges in topic discovery, data collection, and data preparation. International journal of information management, 2018. 39: p. 156-168.

47. Yaqoob, I., et al., WITHDRAWN: Information fusion in social big data: Foundations, state-of-the-art, applications, challenges, and future research directions. International journal of information management, 2016.

48. Jiang, H., et al., A methodology of integrating affective design with defining engineering specifications for product design. International journal of production research, 2015. 53(8): p. 2472-2488.

49. Chan, K.Y., et al., An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness. Journal of engineering design, 2011. 22(8): p. 523-542.

50. Kwong, C.K., H. Jiang, and X.G. Luo, AI-based methodology of integrating affective design, engineering, and marketing for defining design specifications of new products. Engineering applications of artificial intelligence, 2016. 47: p. 49-60.

51. Diego-Mas, J.A. and J. Alcaide-Marzal, Single users' affective responses models for product form design. International journal of industrial ergonomics, 2016. 53: p. 102-114.

52. Barnes, C. and S.P. Lillford, Decision support for the design of affective products. Journal of engineering design, 2009. 20(5): p. 477-492.

53. Kansei/Affective Engineering and History of Kansei/Affective Engineering in the World. 2011, CRC Press. p. 15-26.

54. Lin, M.-B. and J. Wang, A Kansei Engineering Approach to Evaluate Consumer Perception on Social Media: A Case Study of Giant Manufacturing Company.

55. Guo, Y., et al., Improved Fuzzy-Based Taguchi Method for Multi-Objective Optimization of Direct-Drive Permanent Magnet Synchronous Motors. IEEE transactions on magnetics, 2019. 55(6): p. 1-4.

56. Chan, K.Y. and U. Engelke, Varying Spread Fuzzy Regression for Affective Quality Estimation. IEEE transactions on fuzzy systems, 2017. 25(3): p. 594-613.

57. Min, X., B. Yu, and F. Wang, Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD. Sci Rep, 2019. 9(1): p. 2362-2362.

58. Vieira, S., et al., Using Machine Learning and Structural Neuroimaging to Detect First Episode Psychosis: Reconsidering the Evidence. Schizophr Bull, 2020. 46(1): p. 17-26.

59. Silahtaroğlu, G. and N. Yılmaztürk, Data analysis in health and big data: A machine learning medical diagnosis model based on patients' complaints. Communications in statistics. Theory and methods, 2019: p. 1-10.

60. Hong, W.S., A.D. Haimovich, and R.A. Taylor, Predicting hospital admission at emergency department triage using machine learning. PLoS One, 2018. 13(7): p. e0201016-e0201016.

61. Obermeyer, Z. and E.J. Emanuel, Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med, 2016. 375(13): p. 1216-1219.

62. Fleming, A., et al., Missed opportunity? Framing actions around co-benefits for carbon mitigation in Australian agriculture. Land Use Policy, 2019. 85: p. 230-238.

63. Wright, D.W., Civic Agriculture: Reconnecting Farm, Food, and Community. Rural Sociology, 2005. 70(2): p. 276.

64. Lyson, T.A., Civic agriculture: Reconnecting farm, food, and community. 2012: UPNE.

65. Vasquez, A., et al., Community-Supported Agriculture as a Dietary and Health Improvement Strategy: A Narrative Review. J. Acad. Nutr. Diet., 2017. 117(1): p. 83-94.

66. Delgado, J., et al., Big Data Analysis for Sustainable Agriculture. Frontiers in Sustainable Food Systems, 2019. 3: p. 54.

67. McHenry, M.P., Agricultural bio-char production, renewable energy generation and farm carbon sequestration in Western Australia: Certainty, uncertainty and risk. Agriculture, ecosystems & environment, 2009. 129(1-3): p. 1-7.

68. Dutta, P., et al., Blockchain technology in supply chain operations: Applications, challenges and research opportunities. Transportation research. Part E, Logistics and transportation review, 2020. 142: p. 102067-102067.

69. Brewster, C., et al., IoT in agriculture: Designing a Europe-wide large-scale pilot. IEEE communications magazine, 2017. 55(9): p. 26-33.

70. Potts, J., Blockchain in Agriculture. Available at SSRN 3397786, 2019.

71. Al-Jaroodi, J. and N. Mohamed. Industrial applications of blockchain. in 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). 2019. IEEE.

72. Cheung, K.-F., M.G.H. Bell, and J. Bhattacharjya, Cybersecurity in logistics and supply chain management: An overview and future research directions. Transportation research. Part E, Logistics and transportation review, 2021. 146: p. 102217.

73. Cone, C.A. and A. Myhre, Community-supported agriculture : A sustainable alternative to industrial agriculture : Food and power : Case studies in industrial agriculture and its alternatives in the United States. Human organization, 2000. 59(2): p. 187-197.

74. Watson, D.J., Working the fields: The organization of labour in community supported agriculture. Organization (London, England), 2020. 27(2): p. 291-313.

75. Schnell, S.M., Food With a Farmer's Face: Community-Supported Agriculture in the United States. Geographical Review, 2007. 97(4): p. 550-564.

76. Junhong, C., et al., Factors Affecting the Dynamics of Community Supported Agriculture (CSA) Membership. Sustainability (Basel, Switzerland), 2019. 11(15): p. 4170.

77. Salah, K., et al., Blockchain for AI: review and open research challenges. IEEE Access, 2019. 7: p. 10127-10149.

78. Ravindran, A., Operations research methodologies / edited by A. Ravi Ravindran. 2009, Boca Raton: Boca Raton : CRC Press.

79. Rathnayaka, A.J.D., et al., Framework to manage multiple goals in community-based energy sharing network in smart grid. International journal of electrical power & energy systems, 2015. 73: p. 615-624.

80. Multiple Criteria Decision Making. Multiple Criteria Decision Making.

81. McDonagh, D., A. Bruseberg, and C. Haslam, Visual product evaluation: exploring users' emotional relationships with products. Appl Ergon, 2002. 33(3): p. 231-240.

82. Zhang, P. and N. Li, The importance of affective quality. Communications of the ACM, 2005. 48(9): p. 105-108.

83. Chan, K.Y., et al., Affective design using machine learning: a survey and its prospect of conjoining big data. International journal of computer integrated manufacturing, 2020. 33(7): p. 645-669.

84. Linear Regression Analysis with Fuzzy Model. IEEE transactions on systems, man, and cybernetics, 1982. 12(6): p. 903-907.

85. Chukhrova, N. and A. Johannssen, Fuzzy regression analysis: Systematic review and bibliography. Applied soft computing, 2019. 84: p. 105708.

86. D'Urso, P. and T. Gastaldi, An "orderwise" polynomial regression procedure for fuzzy data. Fuzzy Sets and Systems, 2002. 130(1): p. 1-19.

87. Nazari-Shirkouhi, S. and A. Keramati, Modeling customer satisfaction with new product design using a flexible fuzzy regression-data envelopment analysis algorithm. Applied mathematical modelling, 2017. 50: p. 755-771.

88. Yakubu, H. and C.K. Kwong, Multigene Genetic Programming Based Fuzzy Regression for Modelling Customer Satisfaction Based on Online Reviews. 2019, IEEE. p. 1541-1545.

89. Lin, Y.C. and C.C. Wei, A hybrid consumer-oriented model for product affective design: An aspect of visual ergonomics. Human factors and ergonomics in manufacturing & service industries, 2017. 27(1): p. 17-29.

90. Trevor van, G. and A. Edie, Chapter 1 - Why Design for Emotion? 2012, Elsevier Inc. p. 1-18.

91. Breur, T., Data analysis across various media: Data fusion, direct marketing, clickstream data and social media. Journal of direct, data and digital marketing practice, 2011. 13(2): p. 95-105.

92. Fan, W. and M. Gordon, The power of social media analytics. Communications of the ACM, 2014. 57(6): p. 74-81.

93. Jussila, J., et al. Social media analytics empowering marketing insight: A framework for analyzing affective experiences from social media content. in EMAC2016 Conference Proceedings. 2016.

94. Grassi, M., et al., Sentic Web: A New Paradigm for Managing Social Media Affective Information. Cognitive computation, 2011. 3(3): p. 480-489.

95. Cambria, E., Affective Computing and Sentiment Analysis. IEEE intelligent systems, 2016. 31(2): p. 102-107.

96. Guo, K., Y. Tang, and P. Zhang, CSF: Crowdsourcing semantic fusion for heterogeneous media big data in the internet of things. Information fusion, 2017. 37: p. 77-85.

97. Dong, X.L. and D. Srivastava. Big data integration. in 2013 IEEE 29th international conference on data engineering (ICDE). 2013. IEEE.

98. Kwong, C.K., et al., A Modified Dynamic Evolving Neural-Fuzzy Approach to Modeling Customer Satisfaction for Affective Design. ScientificWorldJournal, 2013. 2013: p. 636948-11.

99. Covarrubias, M. and M. Bordegoni, Interaction with virtual aesthetic shapes through a desktop mechatronic system: A desktop system that supports designers in the creation, evaluation and modification of aesthetic virtual shapes. Virtual and physical prototyping, 2014. 9(1): p. 27-43.

100. Kit Yan, C., et al., A Stepwise-Based Fuzzy Regression Procedure for Developing Customer Preference Models in New Product Development. IEEE transactions on fuzzy systems, 2015. 23(5): p. 1728-1745.

101. Kwong, C.K., et al., A generalised fuzzy least-squares regression approach to modelling relationships in QFD. Journal of engineering design, 2010. 21(5): p. 601-613.

102. Lee, S., The role of usability, aesthetics, and cognitive style on user preference for interactive applications. 2010, ProQuest Dissertations Publishing.

103. Fong, A.C.M., et al., Generation of Personalized Ontology Based on Consumer Emotion and Behavior Analysis. IEEE transactions on affective computing, 2012. 3(2): p. 152-164.

104. Chou, J.-R., A Kansei evaluation approach based on the technique of computing with words. Advanced engineering informatics, 2016. 30(1): p. 1-15.

105. Yadav, H.C., et al., Prioritization of aesthetic attributes of car profile. International journal of industrial ergonomics, 2013. 43(4): p. 296-303.

106. Suzianti, A., S. Apriliandary, and N.P. Poetri, Affective Design with Kansei Mining: An Empirical Study from Automotive Industry in Indonesia. 2016, Cham: Springer International Publishing: Cham. p. 76-85.

107. Shen, H.-C. and K.-C. Wang, Affective product form design using fuzzy Kansei engineering and creativity. Journal of ambient intelligence and humanized computing, 2016. 7(6): p. 875-888.

108. Kwon, O., C.-R. Kim, and G. Kim, Factors affecting the intensity of emotional expressions in mobile communications. Online information review, 2013. 37(1): p. 114-131.

109. Kao, T.-Y., et al., Co-creating value with consumers through social media. The Journal of services marketing, 2016. 30(2): p. 141-151.

110. Kwong, C.K., et al., The Hybrid Fuzzy Least-Squares Regression Approach to Modeling Manufacturing Processes. IEEE transactions on fuzzy systems, 2008. 16(3): p. 644-651.

111. Cho, Y. and S. Kim. A Study on Affective Design by Subjective-Objective Co-Approach. in Proceedings of the International MultiConference of Engineers and Computer Scientists. 2014.

112. Chan, K.Y., et al., A Flexible Fuzzy Regression Method for Addressing Nonlinear Uncertainty on Aesthetic Quality Assessments. 2017.

113. Tanaka, H., S. Uejima, and K. Asai, LINEAR REGRESSION ANALYSIS WITH FUZZY MODEL. IEEE Transactions on Systems, Man and Cybernetics, 1982. 12(6): p. 903-907.

114. Dash, S., et al., Big data in healthcare: management, analysis and future prospects. Journal of big data, 2019. 6(1): p. 1-25.

115. Dimitrov, D.V., Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res, 2016. 22(3): p. 156-163.

116. White, S., A review of big data in health care: challenges and opportunities. Open access bioinformatics, 2014. 6: p. 13.

117. Char, D.S.M.D., N.H.M.B.B.S.P. Shah, and D.P. Magnus, Implementing Machine Learning in Health Care -- Addressing Ethical Challenges. The New England Journal of Medicine, 2018. 378(11): p. 981-983.

118. Biller-Andorno, N. and A. Biller, Algorithm-Aided Prediction of Patient Preferences — An Ethics Sneak Peek. The New England journal of medicine, 2019. 381(15): p. 1480-1485.

119. Ahmad, M., et al., Fairness in Machine Learning for Healthcare, in International Conference on Knowledge Discovery and Data Mining. 2020, ACM. p. 3529-3530.

120. Chang, A.C. and J. Hunt, Toward unreasonable effectiveness of cardiac ICU data: artificial intelligence in pediatric cardiac intensive care. Pediatr Crit Care Med, 2014. 15(6): p. 565-567.

121. Roscher, R., et al., Explainable Machine Learning for Scientific Insights and Discoveries. IEEE access, 2020. 8: p. 42200-42216.

122. Blomberg, S.N., et al., 23 Machine learning model surpassing medical dispatchers recognition of out-of-hospital cardiac arrest. BMJ open, 2018. 8(Suppl 1): p. A9.

123. Huertas-Fernández, I., et al., Machine learning models for the differential diagnosis of vascular parkinsonism and Parkinson’s disease using [123I]FP-CIT SPECT. Eur J Nucl Med Mol Imaging, 2015. 42(1): p. 112-119.

124. Mancini, A., et al., Machine learning models predicting multidrug resistant urinary tract infections using “DsaaS”. BMC bioinformatics, 2020. 21(S10): p. 347-347.

125. McCradden, M.D., et al., Ethical limitations of algorithmic fairness solutions in health care machine learning. The Lancet. Digital health, 2020. 2(5): p. e221-e223.

126. Char, D.S., N.H. Shah, and D. Magnus, Implementing Machine Learning in Health Care — Addressing Ethical Challenges. N Engl J Med, 2018. 378(11): p. 981-983.

127. Myles, A.J., et al., An introduction to decision tree modeling. Journal of chemometrics, 2004. 18(6): p. 275-285.

128. Lin, W., et al., An ensemble random forest algorithm for insurance big data analysis. Ieee access, 2017. 5: p. 16568-16575.

129. Natekin, A. and A. Knoll, Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 2013. 7: p. 21.

130. Jakkula, V., Tutorial on support vector machine (svm). School of EECS, Washington State University, 2006. 37.