ANALYTICAL METHODS FOR EFFECTIVE OPERATION OF HUNGER-RELIEF LOGISTICS

by

Rahul Srinivas Sucharitha

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Seokcheon Lee, Chair

School of Industrial Engineering

Dr. Yuehwern Yih

School of Industrial Engineering

Dr. David Johnson

School of Industrial Engineering

Dr. Kee-Hong Kim

Department of Food Sciences

Approved by:

Dr. Abhijit Deshmukh

In dedication to my loving parents for raising me to be the person I am, my brother for his support, my grandparents in Chennai for their guidance, and the Almighty for giving me the strength and courage to achieve my dreams and move forward in life!

ACKNOWLEDGMENTS

This dissertation could not be accomplished without the help, support, and guidance of many people during my Ph.D. study. First and foremost, I would like to deliver my greatest appreciation and sincere gratitude to my advisor, Dr. Seokcheon Lee. Without his academical, financial and personal supports, this work could not be completed. His encouragement and valued advice inspired me to explore new knowledge areas of my interest. I would also like to thank my committee members, Dr. Yuehwern Yih, Dr. David Johnson, and Dr. Kee-hong Kim for their guidance and supervision for this work.

I am grateful to Mr. Dave Kotterman for providing funding resource through Purdue University Technical Assistance Program (TAP) for my Ph.D. study, Ms. Anita Park for helping me handle administrative tasks and the staff members in the School of Industrial Engineering at Purdue University for providing all facilities on campus which enhanced my learning experience and helped me reach the fullest potential in every aspect.

I would like to thank the support received from Food Finders Food Bank, Lafayette, Indiana and Greater Cleveland Food Bank, Cleveland, Ohio. I would like to express my sincere appreciation to Ms. Katy Bunder, the executive director at Food Finders Food Bank and Mr. Phil Trimble, the senior manager of research and program evaluation at Greater Cleveland Food Bank for their contribution towards this research.

I also appreciate my colleagues in the Distributed Control Lab (DC Lab), Dr. Sungbum, Dr. Chul-Hun, Dr. Patchara, Mr.Ho-Young, Dr. Ashutosh, Dr. Yang, Mr. Zekun, Ms. Cansu, Mr. Francisco, Mr. Ibrahim, Mr. Juan, Mr. Yonggab, and many others who always helped me to resolve academic and personal issues.

I would like to thank my entire family for their love, unconditional support and above all believing in me. I am extremely grateful to my wonderful mom, who despite being so far away, never let me feel that I am away from home. She always listened to my daily stories patiently and passionately. Mom – without your unconditional love, care, motivation, and support this would not have been possible. My deepest gratitude goes to my entire family (Sucharitha, Dr. Srinivas, Dr. Atul, Subbarayalu and Renuka) for their support and encouragement to pursue my education in Purdue. I would like to thank Mr. Alexander Struck Jannini for his support and providing invaluable advice and for always seeming to have the best answers to all the hard questions that life and/or research threw at me.

I would also like to thank each and every one of my friends here in the U.S and abroad who have always supported me and provided me strength to move forward in difficult times. Special thanks to Dr. Aishwarya Bhargav, Dr. Rahul Ramamurthy, Dr. Anand Samuel, Mr. Akash Agarwal, Mr. Vignesha Bhamidi, Mr. Abhinav Jayakumar, Mr. Abhishek Sethu, Mr. Kumaraguru Sivasankaran, Mr. Jeff Henline, Ms. Chandnee, Ms. Sharon Rebecca, Mr. Varun Adithya, Mr. Rohith Giridhar, Mr. Atreya Nittala, Mr. Aditya Nittala, Mr. Shravan Raghu, and Mr. Ameya Krishna for their encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Food banks and other non-profit organizations play an essential part in alleviating poverty and improving food security in many countries worldwide. These groups help those in need by providing food and resources. Food banks rely on infrequent food and cash contributions to help them achieve their goals. Due to the finite availability of resources and the dynamic structure, managing food supply and demand in a profoundly uncertain situation like this is difficult. This study tackles these issues by presenting and analyzing various statistical and quantitative models to help food insecure people get food in sustainable and meaningful ways.

The aim and objective of this chapter is to develop and implement data-driven models and analytical techniques, as well as decision support frameworks, to help food bank administrators better understand the dynamics of food donation supply and demand and to improve the accuracy of food supply and demand behavior prediction at various planning levels to ensure equitable and efficient distribution of food.

First, a systematic review was done to research the evolving literature in food bank logistics. A perusal of the literature shows that research in food bank logistics is evolving, and issues about fairness, sustainability, cost reduction, food quality and nutrition, data uncertainty, and food waste study have not been reviewed in great detail. This study attempts to fill this existing gap utilizing a literature review on these issues and outline future research directions based on research gap analysis. Forty-eight published articles were selected, categorized, analyzed, and literature gaps were identified to suggest future research opportunities. The review will provide its usefulness for academicians, researchers, and experts to better understand food bank logistics and guidance for future research.

Second, a unique framework of a hybrid model combining ARIMA and neural network autoregressive (NNAR) model to capture both linearity and nonlinearity in the univariate analysis of the food donation supply is proposed. We introduce an iterative cross-validation method called walk-forward cross-validation to the hybrid methodology. Each model’s parameters can be varied and tested again on an iterative basis to obtain optimized tuning parameters specific to the

algorithm. The proposed hybrid approach and methodology are applied to the food supply datasets and give better forecasting accuracy than the state-of-the-art. Additionally, the food supply behavior study is further extended for a multi-variate analysis by leveraging statistical and machine learning algorithms to identify the key predictors of the food supply behavior using the same historical food supply data of a regional food bank. Based on the numerical study, Random Forest (RF) method best captures the complex structure of the data compared to the other developed predictive models. Furthermore, we provide a useful framework for implementing these models for their effectiveness in a non-profit organization such as the food-aid distribution system and implementing the proposed framework for several use case studies based on different levels of planning to noteworthy ease and comfort intended for the respective planning team.

Thirdly, understanding the dynamics of the demand that food banks get from food insecurity has significant importance in optimizing operational costs and equitable distribution of food, especially when demand is uncertain. Hence, Gaussian Mixture Model (GMM) clustering is selected to extract patterns. The novelty is that GMM clustering is applied to identify the possible causes of food insecurity in a given region understanding the characteristics and structure of the food assistance network in a particular region, and the clustering result is further utilized to explore the patterns of uncertain food demand behavior and its significant importance in inventory management and redistribution of surplus food thereby developing a two-stage hybrid food demand estimation model. A food bank network in Cleveland, Ohio is used as a case study and the clusters developed are studied and visualized. The results reveal that this proposed framework can make an in-depth identification of food accessibility and assistance patterns and provides better prediction accuracies of the leveraged statistical and machine learning algorithms by utilizing the GMM clustering results.

Finally, the analytical methods implemented and developed to study the supply and demand of the food donations are extracted and used to develop a conceptual framework for designing a decision support system to apply visual analytics to a food bank's decision-making process. To validate the conceptual framework, a decision support system has been designed accordingly, and dashboards are developed to improve a food bank's strategic, tactical and operational planning. The findings of this research can help food banks make better decisions and manage the resources efficiently and serve the people in need. It also has the potential to be further expanded to other hunger-relief organizations.

1. INTRODUCTION

Food insecurity is defined as the condition of limited availability of cost-effective, accessible, and nourishing food in socially acceptable ways (Alotaik et al., 2017; Hampl & Hall, 2002; Mousa & Freeland-Graves, 2017). According to the 2014 insights provided by Feeding America - one of the largest hunger-relief networks in the United States, 42.2 million of the total population survive in food-insecure households, including 29.1 million adults and 13.1 million children. This comprises about 13.1% of food insecure families (Feeding America, 2014), and the numbers continue to grow. As shown in Figure 1.1, the number of people in the United States who are food insecure had been on the rise until 2008, remaining stable for a certain period and declining slightly from 2011 to 2013. This issue is being faced by several countries (both developed and developing) all over the world. To control this issue, food-insecure people and families obtain assistance from the government and food rescue organizations. One such food rescue program in the United States is Feeding America. Feeding America distributes food and aid to these individuals by having a nationwide network of around 200 food banks and around 60,000 food pantries and meal programs. As depicted in Figure 1.2, food rescue and distribution in the United States involve the management of food, material, currency, and other sources that aid food delivery to the people in need. The food banks and other non-profit organizations play a pivotal role in collecting and distributing foods to the food insecure, either directly or via agencies that serve local communities (Feeding America, n.d.). These agencies include all the food pantries and meal programs that the food banks support. Besides managing the collection of donated foods from the various food donors, food banks also manage the distribution of these donated foods to the various food agencies associated with it. The research problem in this research is motivated by the scheduling challenges faced by the food bank distribution chain.

.

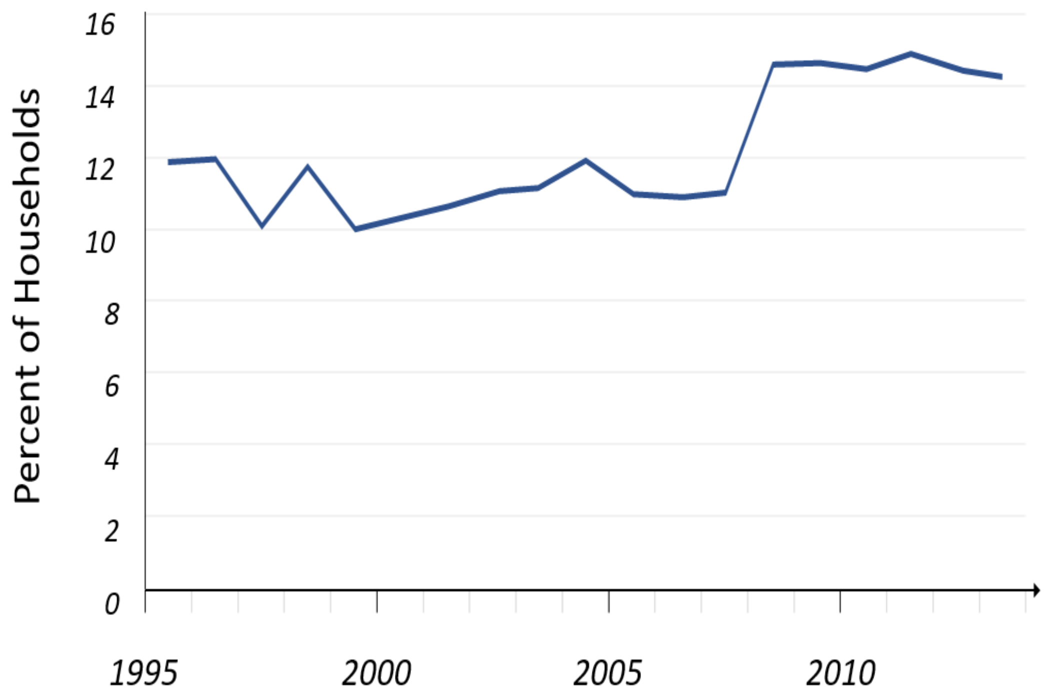


Figure 1.1 Food insecurity percentages in the U.S., 1995-2013. (Sengul Orgut et al., 2016)

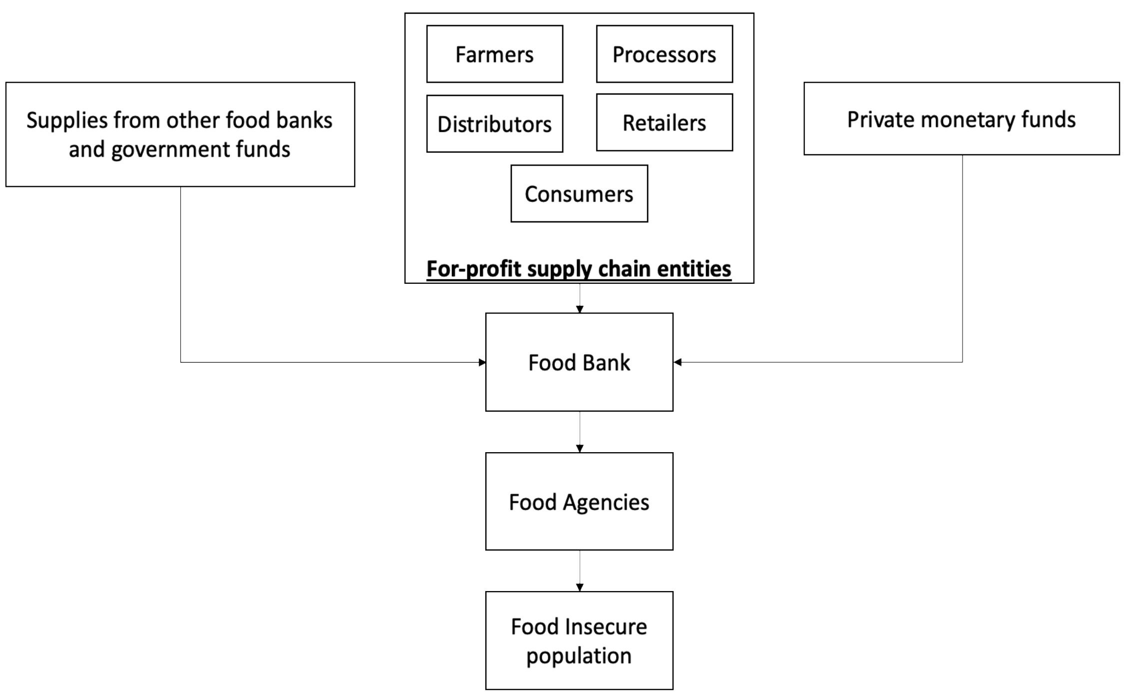


Figure 1.2 Management of resources in the food relief supply chain

Food banks are autonomous in their operations (Brock & Davis, 2015; Lien et al., 2014; Orgut et al., 2016b) with the exception that they report back to Feeding America regarding the amount of donated food and services they have distributed in their local community considering the size of the food insecure population in that given region. Food banks obtain donated food and grocery products from food donors such as national food and grocery manufacturers, vendors, etc., within scheduled due dates and time windows. The donated food is carried back to the food bank using either rented or owned trucks. These donated foods are checked in the food bank for quality reasons. The trucks are subsequently used to deliver the donated foods to numerous food pantries and meal programs (also termed as food agencies) based on their accessibility and requested service times. The trucks return to the food bank once the donated foods are delivered. The individuals receive the donated food items from these food agencies in their region.

Management of the food bank supply chain is a prominent problem considering that the supply chain involves the movement of donated food that must be delivered to the food insecure people quickly. If not, there would be a significant amount of food waste and hence an increase in food malnutrition, considering that food waste and malnutrition have a close relationship (Parfitt et al., 2010). The said challenge points to the need for food banks to have a flexible distribution process. Apart from distributing the donated foods to various food agencies, as many of these agencies lack the resources to obtain food regularly, food banks also make the food available in their organization or warehouse to serve those close to the food bank. It is constantly being taken into consideration by the food banks to increase the accessibility of the donated foods to those agencies that are in remote areas. To accomplish that, there are fixed weekly schedules of transportation being carried out by the food banks to these areas to make sure that the demands are met.

A non-profit organization like the food banks does not have profit as their only objective. The main goal is to ensure that food insecurity is reduced, and to do that, efficient usage of the available resources is critical.

1.1 Research Statement and Contributions

Food banks depend on food and cash donations that infrequently occur to help them fulfill their goals. In highly uncertain conditions such as this, balancing the supply and demand of food is challenging considering the limited availability of resources and the complex system. This research addresses these challenges and presents and analyses several statistical and mathematical models to facilitate food distribution to the food insecure sustainably and effectively. The contributions of this dissertation are:

- Fill the existing gap in the literature by conducting a systematic literature review and identify the current issues for research based on research gap analysis
- Develop a novel hybrid time series model combining ARIMA and neural network autoregressive (NNAR) model for a univariate analysis of food bank supply
- Conduct a comprehensive numerical study and develop a multi-variate analysis to quantify the extent of uncertainty in terms of donor and funding sources and their characteristics, food type and storage type using a novel predictive framework
- Develop a two-stage hybrid food demand estimation model using Gaussian Mixture Model (GMM) clustering Bayesian Additive Regression Trees (BART)
- Develop an interactive dashboard to assist the decision-making process using a novel conceptual framework

1.2 Organization of Dissertation

Each chapter of this dissertation has its introduction and relevant background associated with a problem or class of problem that a decision-maker might face. The remaining of the chapters of this dissertation are organized as following sequences. In Chapter 2, the systematic literature relevant to food bank logistics is reviewed. Chapter 3 presents the hybrid time series methodology for the food supply dataset. The multi-variate analysis of the food supply dataset is presented in Chapter 4. In Chapter 5, a two-stage hybrid food demand estimation model is explained, and in Chapter 6, the conceptual framework for designing a decision support system to apply visual analytics to a food bank's decision-making process is presented. Finally, Chapter 7 concludes this dissertation and summarizes the plan for further research directions.

2. COMPREHENSIVE LITERATURE REVIEW AND PERSPECTIVES IN FOOD BANK LOGISTICS

When a person is hungry or is in a situation of hunger, it is a feeling that is felt physically, but in terms of the intensity of hunger, subjectivity arises and hence, challenging to gauge or quantify. Currently, the means of measuring hunger and food inadequacy is done for a sizeable amount of people or a large population. The term used is food insecurity. The description of food insecurity keeps changing overtime depending mainly upon the region in which it is measured and the perspectives of hunger in each region, thereby widening the research interest of food insecurity. Food insecurity is the circumstance or situation of scarce availability of affordable and nutritious food in socially customary ways (Alotaik et al., 2017; Hampl & Hall, 2002; Mousa & Freeland-Graves, 2017). In the United States, around 50 million people face hunger. This constitutes about 1 in 6 of the country's population, encompassing more than 1 in 5 children facing hunger as well (Feeding America, 2014). Food supply chain literature has evolved over the years as not just focusing on the profit-based food supply chains, but also taking into account the hunger-relief logistics, thereby extending the literature to include long-term humanitarian issues such as food insecurity, specifically non-profit food supply chains as an extension for research interests.

Food insecurity affects about 13.1 percent of households, and the figure continues to rise (Feeding America, 2014). According to Feeding America, one of the largest domestic hunger-relief organization in the United States, 42.2 million Americans, including 29.1 million adults and 13.1 million children, live in food-insecure households. Food insecurity in the United States was on the increase until 2008 but remained stable for a while before decreasing slightly from 2011 to 2013 and will continue to rise as a result of the current pandemic crisis (Providence, 2020). This issue is being faced by several countries (both developed and developing) all over the world (Campbell, 1991; Casey et al., 2001; Lee & Frongillo, 2001; Nelson et al., 2001; Kumar et al., 2009; Vozoris & Tarasuk, 2003) have addressed the impact of food insecurity on the quality of life and health status of individuals. There are government and food rescue systems to combat the issue of food insecurity in America. Food insecure individuals and households receive support through a nationwide network of about 200 food banks and 60,000 food pantries and soup kitchens, handled and managed by Feeding America – the leading food rescue service in the United States. The

Feeding America network serves over 37 million people per year, comprising of 14 million children and 3 million seniors. Feeding America's clients face situations where they had to choose between paying for food and paying for medications or medical attention in 34% of cases. Furthermore, 72 % of food banks believe they would not be able to fulfill the demands of their families without changing the volume of food supplied (Feeding America, 2014).

Based on the work carried out by (Orgut et al., 2016b), different key stakeholders and stages have been established and are depicted in Figure 2.1. Food banks and other non-profit organizations play a critical part in gathering and providing food to those in need, either directly or by community-based agencies (Feeding America, 2015). These agencies include all of the food bank’s supported food pantries and meal programs. The food bank supply chain network involves the handling of food, materials, currency, and other resources that help in the supply of food to the food insecure. Food banks play a critical part in reclaiming excess food that would otherwise be thrown away. Big farms and food processors have some harvested crop left over which the food banks store. Defective foods, which may or may not be sellable, are often purchased from food producers and dealers. However, the bulk of the food supplies come from grocery stores and food producers. Food banks oversee the procurement of donated foods from various food donors as well as the delivery of these donated foods to the various food agencies with which they are affiliated.

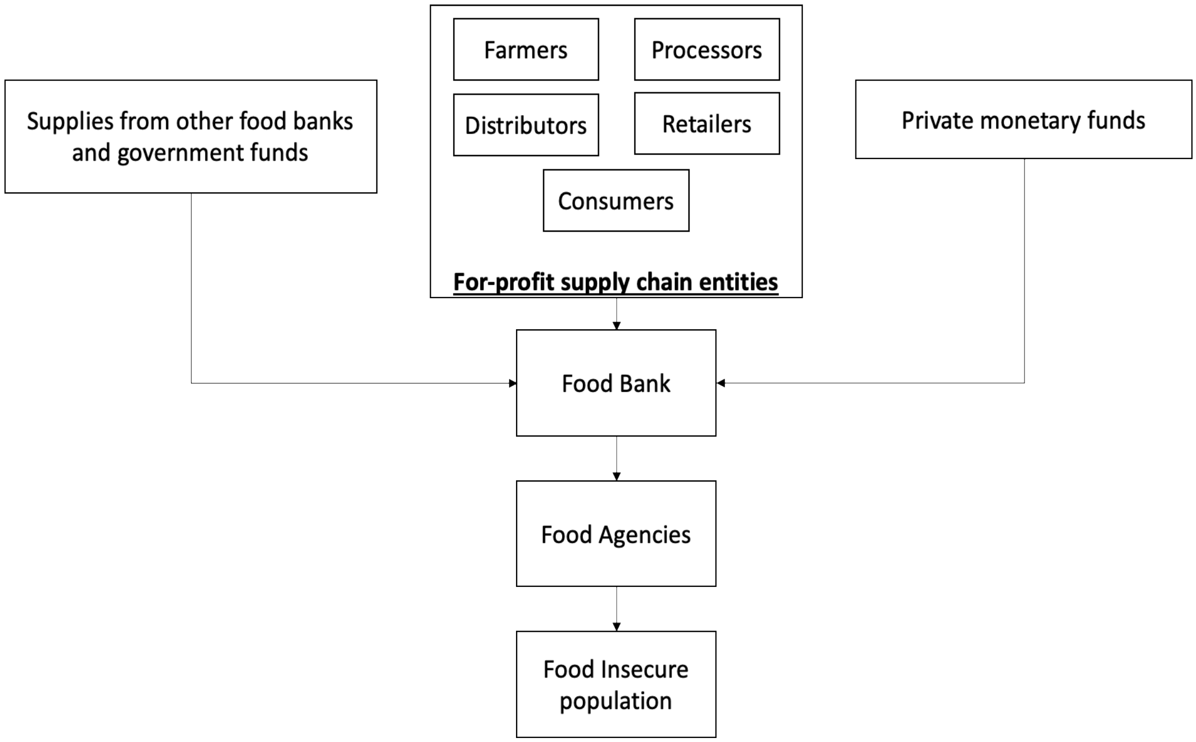


Figure 2.1 Management of resources in the food relief supply chain

Food banks are autonomous in their operations (Davis et al., 2014; Lien et al., 2014; Orgut et al., 2016b) with the exception that they report back to Feeding America regarding the amount of donated food and services they have distributed in their local community considering the size of the food insecure population in that given region. Many scholars have explained the delivery of donated foods from various viewpoints. Food banks receive donated food and grocery supplies from food lenders such as major food and grocery distributors, merchants, and others, under predetermined deadlines and time frames. Using borrowed or owned trailers, the donated food is returned to the food bank. These donated foods are stored at the food bank to ensure that they are of high quality. The donated foods are then distributed to a number of food pantries and meal services (known as food agencies) depending on their availability and requested service hours. When the donated goods have been delivered, the trucks head to the food bank. People in their area collect donated foods from local food pantries and meal services. Aside from food donations, Feeding America foodbanks invest in the ChoiceTM scheme, which was created in collaboration with the University of Chicago (Prendergast, 2017). The Choice system is an online/mobile-enabled auction system that allows food banks to place orders for food depending on the number of shares they have. These organizations can also rely on food banks to provide extra food allotments based on donations received locally. Food banks are also engaged in emergency recovery, supplying millions of pounds of food to people in disaster-stricken countries (Feeding America, 2015). Food banks accept cash contributions in addition to food donations to help with procurement and logistics. These private monetary funds are used by food banks to buy goods that are in high demand, have a long shelf life, and are reasonably priced. For a limited portion of the procurement price, these acquired items are made available to their food agencies. As a result, products purchased by private monetary contributions differ from those obtained through food donations.

The literature on foodbank logistics is limited. To our knowledge, no comprehensive literature review has been conducted that focuses on food bank delivery and collection modeling, as well as a comparative analysis of food bank logistics. (Bazerghi et al., 2016) and (McKay et al., 2019) looked at 35 and 57 studies on the role of food banks in maintaining food security, respectively. Food banks have urgent remedies to food insecurity, but they lack the ability to maintain overall food security due to a lack of nutritionally based food banking initiatives in the past, according to these reviews. These articles have concentrated on gathering qualitative research from the food bank literature rather than quantitative research or the technical facets of the food bank network. An et al. (2019) supplemented this analysis by analyzing 14 articles and concluding that providing a diet-focused donated food delivery scheme for the food deprived improves the long-term success of food pantries.

We thoroughly reviewed the literature and heeded the need for study by answering the following research questions in order to shed some light on how to improve identification of logistical problems and issues faced by food banks and their food agencies:

- RQ1. What are the key elements, logistical aims and concerns, of the food bank supply chain in the literature? What is the state of research on such elements on a quantitative level of the outcome?
- RQ2. What research gaps can guide future studies?

The proposed article aims at introducing and filling the huge gap of studying the quantitative models, logistical aims, and modeling challenges in the food bank literature and finding research gaps for identifying the future scope of research.

This study makes an attempt to present a comprehensive review of the published literature on the logistical issues faced by food banks. Subsequently, the study analyzed the research gaps in the literature to facilitate further study, and research directions. The remaining part of the study is organized as follows: Section 2.1 defines the Research methodology. Detailed discussions on specified classifications are undertaken in Section 2.2. Section 2.3 discusses and critically analyzes research gaps, and the study is summarized in Section 2.4 by offering conclusions along with its limitations.

2.1 Research Methodology

The aim of this paper is to systematically analyze the current research in the improvement of the food bank logistics systems and identifying the latest progress in ensuring food security. We accomplish this aim by conducting a systematic literature review utilizing aspects of structured literature study as mentioned by (Denyer et al., 2008). they suggest that the two main purposes of a systematic literature review are to combine all the research findings in a specific area and identifying research gaps that can guide future research. To further enhance the thoroughness of our literature review, we apply an inclusion and exclusion criteria (see Table 2.1). In addition to guiding the research, these criteria also support rigorous and defensible data (Meline, 2006). The method is used to systematically evaluate the themes of recorded information. It is useful for creating thorough literature reviews because it allows for understanding the focus of written text in a rule-governed way, thus enhancing replicability.

This review consists of a comprehensive three-step process for conducting literature reviews. The three steps are- Material collection, descriptive analysis, and category identification and analysis each described in the following section.

2.1.1 Material collection

Material collection methodology and unit of analysis is the first step of the literature review process. The unit of analysis has been defined as a single research article/dissertation/book/report. The Boolean phrase and the inclusion/exclusion criteria used for the study can be seen in Table 2.1. As part of this review process, only peer reviewed articles available in English were considered for inclusion. The articles were collected from Google-scholar search engine (www.scholar.google.com) and in Scopus (www.scopus.com) with options of searching for articles in English Language excluding articles in other languages and sorted by relevance. Lastly, in the search we targeted papers published in the period ranging from January 2013 to June 2020. This point of direction was chosen based on the publishing dates of seminal publications on developing food bank logistics.

Table 2.1 The Boolean phrase used as well as inclusion/exclusion criteria

Boolean Phrase	Inclusion/Exclusion Criteria
(“Food bank*” OR “hunger relief*” OR “food bank supply chain*” OR “food bank supply chain network*”) AND (“nonprofit*” OR “not-for-profit*”)	- Must be peer reviewed - Must be written in English - Articles must be focused on food bank supply chain management and control

A total of 1130 articles results from the keyword search. After removing duplicates and filtering for peer-reviewed impact factor publications, 523 articles remained for evaluation.

We then proceeded to review the abstracts of these 523 articles to assess if they fit our research questions. Accordingly, only articles with the quantitative focus that addressed food bank logistics optimization were considered relevant for further analysis. This reduced the article dataset from 523 to 84 articles considered for further review. The full text of these 84 articles was reviewed in depth to enhance the comprehensiveness of our review and also located additional papers relevant to our review and identified 12 additional articles that were considered relevant for our systematic review in food bank logistics.

Subsequently, 48 articles were chosen for our study that fit our research questions. Figure 2.2 Depicts the PRISMA chart search process.

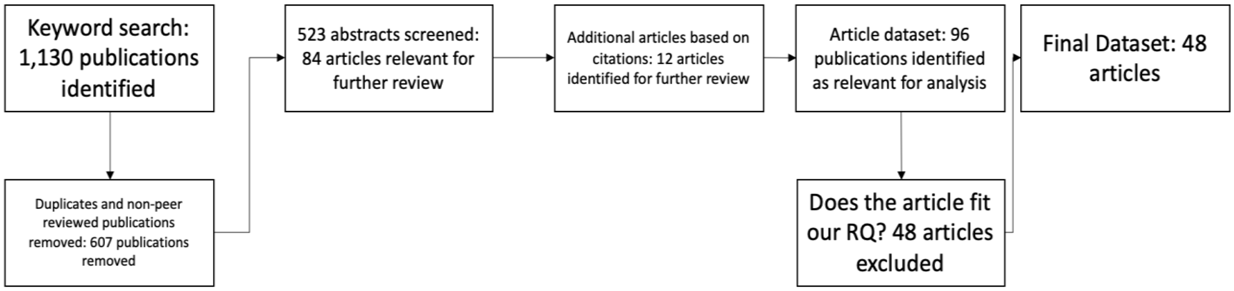


Figure 2.2 Article search, evaluation and exclusion process

2.1.2 Descriptive analysis

In the descriptive analysis step, the formal characteristics of the articles collected are assessed with the aim of providing background for subsequent evaluation of each article's content. The annual distribution of number of articles published for year 2013 to 2020 is shown in Figure 2.3. Most of the articles have been published in reputed journals such as *European Journal of Operational Research, International Journal of Production Research, and IISE Transactions*. It is also clear that numbers of articles have increased considerably in last few years because of growing interest of researchers in this area. Highest number of articles has been published in the year of 2018.

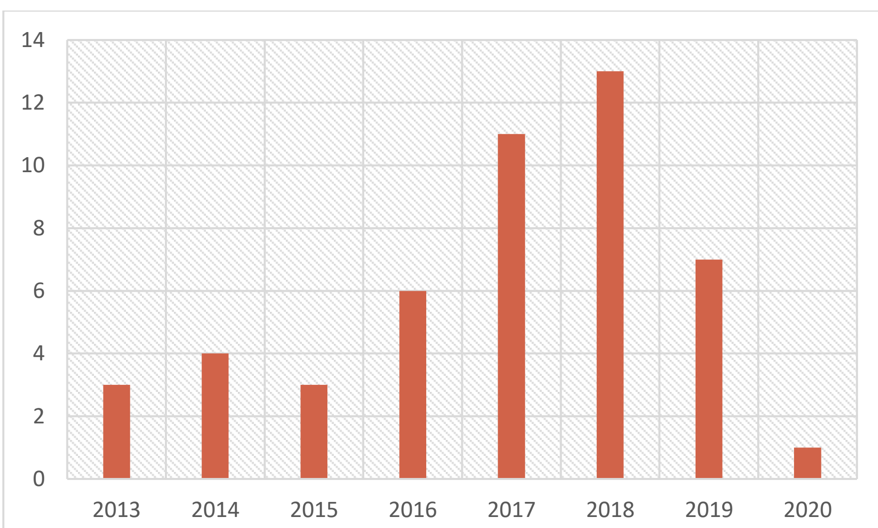


Figure 2.3 Annual distribution of publications across the period of the study

2.1.3 Category identification

Categories of the study and framework for the study are shown in Figure 2.4. As discussed in section 2.1, due to the lack of depth in previous literature, the literatures on food bank logistics are classified into five categories. These five categories are (1) Improved fairness and sustainability;(2) Reduction of uncertainty; (3) Reduction of food waste; (4) Improved food quality and nutrition;(5) Cost reduction. This classification and framework for the study is shown in Figure 2.4.

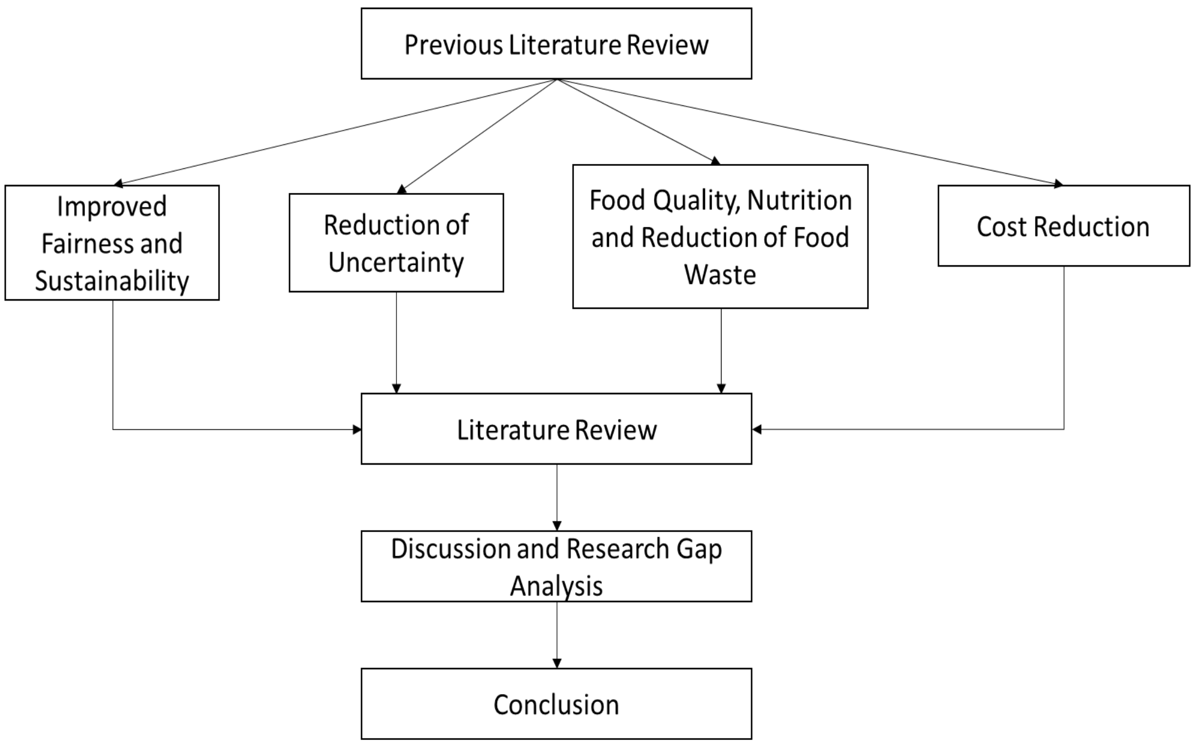


Figure 2.4 Framework of the study

Distribution of research articled for the five categories is shown in Figure 2.5. Cost reduction has maximum percentage of 34% of the publications while most of the other categories are comparatively having lesser percentages of articles indicating the need for future scope of research in these areas. We also consider the drivers and enablers of the categories as key logistical aims to provide the potential research intentions (Table 2.2). Discussing the drivers and enablers, and the explanations of the categories as the key logistical aims, allows us to evaluate and assess respectively the KPIs and the logistics system scope of the models in the further sections. The definitions of each of the KPIs are as follows:

1. Total Logistics Costs incurred (LGC) – it is the sum of all the costs sustained during the flow of products in the supply chain.

2. Storage capacity usage (SCU) -it is the measure of the storage space in a warehouse used and is usually quantified in volume units or percentage of space occupied.

3. Transportation costs (TC) – Costs of transportation of goods and supplies.

4. Investment capital incurred (ICI) – is the amount of money raised by an organization used for undertaking new projects.

5. Completion time of task (CTT) – Time taken to complete a task under observation.

6. Speed of response (SR) – time taken to react to a particular action.

7. Volunteer capacity usage (VCU) – percentage of volunteers from the available database.

8. Volunteer service rate (VSR) – to measure value of volunteers, it is measured by total volunteer hours and/or service attendance.

9. Equitable distribution of food (EDF) – distributing fairly and equally with all concerned generally quantified using user-defined functions and metrics.

10. Carbon emissions (CE) – generally quantified based on the fuel economy and rate of emission by the source in a given supply chain.

11. Supply traceability (ST) – rate of accuracy in predicting the supply of a quantifiable item.

12. Monetary donations obtained (MDO) – amount of capital obtained in the form of donations to the food bank organization.

13. Unsatisfied demand (UD) – percentage of demand not fulfilled by means of food served to the food insecure

14. Demand traceability (DT) - rate of accuracy in predicting the demand of food insecure population.

15. Food insecurity visibility (FIV) – Visualization of the percentage of food insecure in a given region.

16. Nutritional satisfaction (NS) – Measure of availability of nutritious food as per the USDA guidelines in a given food bank organization.

17. Increased shelf-life (IS) – Measure of the freshness of the food item for edible consumption.

18. Disposal costs (DC) – costs incurred in food wastes.

19. Food waste reduction (FWR) – quantified in percentage of food served by the food banks to the food insecure within the given shelf life of the food item.

20. Food donation quality (FDQ) – measure of shelf-life of food donated by food donors to the food bank organization.

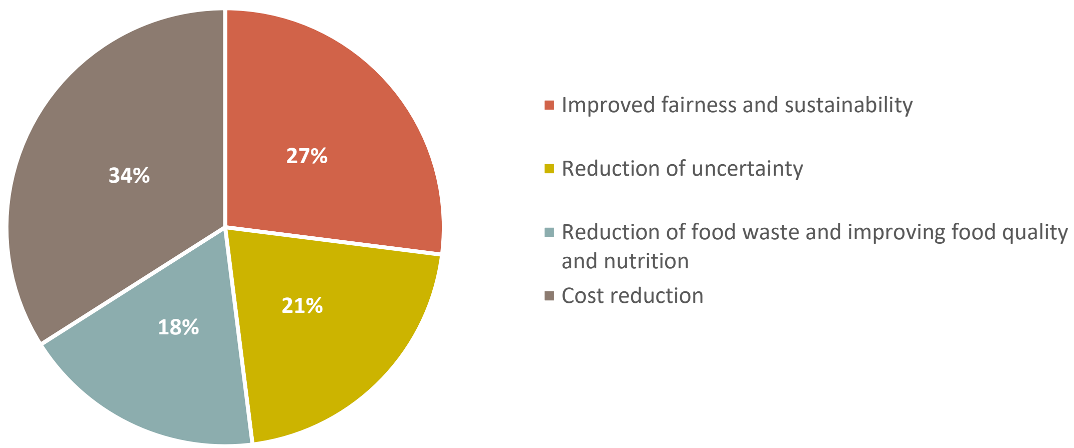


Figure 2.5 Distribution of research papers for different categories

Table 2.2 Key logistical aims in Food Bank Supply Chain

Key Aims	Drivers and Enablers	Explanation of Key Aims
Improved fairness and sustainability	– Growth of Food insecurity – Increase of sustainability awareness – Food scarcity – Social Welfare	The ability to reduce societal impacts related to equitable distribution of food, food insecurity, and environmental concerns of operations
Reduction of uncertainty	– Demand satisfaction – Donated food supply collection – Lack of coordination – Delivery constraints – Managerial implications	The need to have complete visibility of the supply and demand of the donated foods allowing to forecast and study the data sources throughout the stages in the food bank supply chain
Reduction of food waste	– Demand for edible, nutritious food – Products with longer shelf life – Organizational and regulatory solutions – Donor collaborations	The need to collaborate in the food bank supply chain network to reduce food wastes that can be used for food donation because of edible quality
Improved food quality and nutrition	– Increased concerns for food security – Education and awareness raising – Consumer focused awareness raising	The ability to control and process food quality in the donation process and deliver nutritious and edible food to the food insecure by incorporating quality decay information in the logistics decision making
Cost reduction	– Lack of monetary funds – Economic crisis – Volunteering workforce	The need to minimize the total logistical and network costs from the source of the food donations to the final point of donated food consumption

2.2 Detailed analyses of the literature

The selected articles of the literature review are discussed and explored in this section to construct a holistic view of the current studies in Food Bank logistics. The results will illuminate the current gaps and future directions for research. First, we present the main characteristics of the reviewed literature (Table 2.3).

2.2.1 Modeling Characteristics

In recent years Operations Research and logistics literature in the food bank research has shown a growing interest. In this study, we investigate the various logistics and quantitative models with respect to the main characteristics (Table 2.3) summarized below:

Modelling type: Researchers develop various types of models to facilitate the decision-making process to be carried out in a systematic way. The distribution of model types used in the batch of 48 papers are as follows: (i) Mixed Integer Programming (54% of all models), (ii) Analytical (19%), (iii) Simulation (19%), (iv) Linear Programming (6%), (v) Multi Objective Programming (2%).

(Non)linearity: There is a mix of linear and non-linear models in the literature.

Solution approaches and tools: Apart from standard software programs (e.g. Cplex, Lindo), various heuristics have been developed to solve the models due to the complexity of the problem, large problem instances, or possibility to generate fast solutions that lead researchers to consider heuristic approaches.

Real vs. Hypothetical: Proposed models are implemented either by considering real or hypothetical data.

Key Performance Indicators (KPI): A quantifiable value that demonstrates how effectively objectives are achieved.

Table 2.3 Model Characteristics

Sl No.	Articles	Model Type	(Non) Linearity	Solution Approaches and Tools	Real vs. Hypothetical	Key Performance Indicators (KPI)
1	(Jiang et al., 2013)	Analytical	NL	U	R	ST, FWR
2	(Gharehyakheh, 2018)	Analytical	NL	U	R	IS, FWR
3	(Hindle & Vidgen, 2018)	Analytical	NL	U	H	EDF
4	(Desai, 2015)	Analytical	NL	SAS, Tableau	R + H	CTT
5	(Mohan et al., 2013)	Analytical	NL	R	R	EDF
6	(Aiello et al., 2015)	Simulation	NL	Arena	H	SR
7	(Balza-Franco et al., 2017)	LP	L	Microsoft Excel Solver	H	LGC
8	(Nair et al., 2016a)	Simulation	L	U	H	LGC
9	(Yang, 2018)	MIP	L	CPLEX, Heuristic	R + H	TC
10	(Sucharitha & Lee, 2019)	Simulation	NL	Heuristic	R	EDF, FWR
11	(Sucharitha & Lee, 2018)	Analytical	NL	R	R	DT
12	(Ata et al., 2019)	Analytical	NL	VBA	H	EDF, FWR
13	(Rey et al., 2018)	Simulation	NL	Arena	R	VCU, VSR, EDF
14	(Eisenhandler & Tzur 2019a)	MIP	L	ILP Solver, Heuristic	R + H	TC, EDF
15	(Eisenhandler & Tzur 2019b)	MIP	L	ILP Solver, Heuristic	R + H	TC, EDF

Table 2.3 continued

16	(Nair et al., 2016b)	MIP	L	CPLEX, Heuristic	R + H	LGC, EDF
17	(Nair et al., 2017a)	Analytical	NL	U	R	ST, FWR
18	(Ferreira et al., 2018)	MIP	L	CPLEX, Heuristic	H	LGC, EDF
19	(Rey et al., 2018)	MOP	NL	C#	H	LGC, DC
20	(Lee et al., 2017)	MIP	L	CPLEX, Heuristic	R	LGC
21	(González-Torre & Coque, 2016)	SP	NL	Arena	R + H	VCU, EDF
22	(Bacon & Baker, 2017)	Analytical	U	U	R	FWR
23	(Gharehyakheh et al., 2019)	Analytical	NL	GIS	R	FIV
24	(Pollastri et al., 2018)	SP	NL	Arena	R + H	VCU, EDF
25	(Galli et al., 2019)	Simulation	NL	U	R + H	FDQ
26	(Martins et al., 2019)	MIP	L	CPLEX, Heuristics	R + H	LGC, SCU, TC, ICI, EDF, CE, UD
27	(Orgut et al., 2018)	Simulation	NL	U	H	FWR
28	(Schneider & Nurre, 2019)	MIP	NL	U	R + H	LGC, EDF
29	(Ahire & Pekgün, 2018)	MIP	L	CPLEX, Heuristics	R + H	LGC
30	(Brock & Davis, 2015)	Analytical	U	U	R	FWR
31	(Glover et al., 2014)	Analytical	NL	R	R	LGC

Table 2.3 continued

32	(Solak et al., 2014)	Simulation	NL	Genchi G2014enbutsu, TPS tools	R	LGC
33	(Reihaneh & Ghoniem, 2018)	MIP	L	CPLEX, Heuristics	R + H	LGC
34	(Alotaik et al., 2017)	MIP	L	C#, Heuristics	H	LGC
35	(Aboujaoude & Schneider, 2017)	Analytical	NL	U	R	ST
36	(Orgut et al., 2017)	Simulation	U	VBA	R	LGC, VSR
37	(Balcik et al., 2014)	SP	NL	U	R + H	LGC, EDF
38	(Orgut et al., 2016a)	MIP	NL	Heuristic	R + H	LGC, FWR
39	(Davis et al., 2014)	MIP	L	CPLEX, Heuristic	R + H	LGC, EDF
40	(Sönmez et al., 2016)	SP	NL	U	R + H	LGC, EDF
41	(Buisman et al., 2019)	Analytical	NL	GIS	R	LGC, DT
42	(Ortuño & Padilla, 2017)	MIP	L	Xpress-IVE	R	LGC
43	(Prendergast et al., 2017)	MIP	L	Lingo	R	LGC, NS
44	(Nair et al., 2017b)	Simulation	NL	U	R + H	LGC, MDO, FDQ
45	(Davis et al., 2016)	Analytical	NL	U	R	ST, FWR
46	(Ibarra-Rojas & Silva-Soto, 2020)	Analytical	NL	U	R	ST, FWR
47	(Delpish & Jiang, 2019)	Analytical	NL	U	R	SCU, ST
48	(Phillips et al. 2013)	Analytical	NL	U	R	SCU, ST

Table 2.4 Notation

Notation	**Meaning**
MIP	Mixed Integer Programming
LP	Linear Programming
MOP	Multi-Objective Programming
SP	Stochastic Programming
IP	Integer Programming
U	Unspecified
L	Linear
NL	Non-Linear
R	Real
H	Hypothetical
LGC	Total Logistics Costs incurred
SCU	Storage capacity usage
TC	Transportation costs
ICI	Investment capital incurred
CTT	Completion time of task
SR	Speed of response

Table 2.4 continued

VCU	Volunteer capacity usage
VSR	Volunteer service rate
EDF	Equitable distribution of food
CE	Carbon emissions
ST	Supply Traceability
MDO	Monetary donations obtained
UD	Unsatisfied demand
DT	Demand traceability
FIV	Food insecurity visibility
NS	Nutritional satisfaction
IS	Increased shelf-life
DC	Disposal costs
FWR	Food waste reduction
FDQ	Food donation quantity

2.2.2 Improved fairness and sustainability

Tackling hunger and food insecurity are the main aims of foodbanks, and this is done by supplying and distributing food in their service region with the help of donated foods. It is usually the case when the demand is way higher than the supply of donated foods in this logistics system (Orgut et al., 2016a). Due to this situation, the focus of food banks is to provide as much food as possible and as equitably and sustainably possible to the benefactors and the underprivileged. Equity and sustainability stand out in a non-profit organization because of its contradictions with the other objectives that arise in an organization (such as, minimization of wastes, costs, etc.,) (Russell, 2005). Hence, this has been a focus in several research papers in the recent years and because of which, equity and harmful gas emissions has been considered as the main Key Performance Indicators (KPIs) in the study for improving fairness and sustainability as can be seen in Table 2.
Lien et al. (2014) presented a single vehicle sequential resource allocation model for a hunger relief program in Chicago, with the objective of equitable and effective distribution of donated food. An egalitarian welfare utility function is taken into consideration as the study for equity. Balcik et al. (2014) extended this single vehicle sequential resource allocation model to a multi-route setting. Orgut et al. (2016a) designed a joint decision-making model for the fair distribution of storage capacities in various counties of North Carolina to tackle the storage space issues. Fairness and equitability are considered by minimizing the absolute variation between the proportion of food delivered and the demand of the counties and expanding the total food delivered in parallel. Orgut et al. (2017) extended this model to include stochasticity. Orgut et al. (2018) developed robust optimization models to further combat capacity uncertainty and improve the overall equity constraints at all the bottleneck locations. However, these models looked into the inventory management of the network. For the inclusion of the routing costs, Davis et al. (2014) developed transportation schedules for food donation collection and food delivery by instituting food delivery points and implementing a set covering model to assign food agencies to each food delivery point.

Nair et al. (2016a) developed routing models that incorporated three objective functions to report efficiency and fairness in allocation namely- maximize the total utility of the agencies (utilitarian); maximize the utility of the worst-off agency (egalitarian); and minimize the deviation of the utilities of the best- and worst-off agencies. These proposed fairness objectives aid in the fair allocation of the demand that arises in various food agencies of the network. Martins et al. (2019) developed a food bank supply chain network model and addressed the trade-offs and conflicting results that arises with the inclusion of equity and sustainability in a cost reduction focused situation. A mixed integer programming model was proposed with three objective functions that considered economic, environmental, and social dimensions of the network and computational study was conducted to investigate the trade-offs that occur under the three conflicting objectives. Sucharitha and Lee (2018) developed a food distribution policy using suitable welfare and poverty indices and functions to ensure an equitable and fair distribution of donated foods as per the varying demands and requirements of the people. (Yang, 2018; Eisenhandler & Tzur 2019a; Ibarra-Rojas & Silva-Soto, 2020) developed Sequential Resource Allocation problems with the focus of equity maximization and developed algorithms that provides high-quality solutions.

2.2.3 Reduction of Uncertainty

There is considerable relevant literature discussing the role of machine learning and data mining models in the improvement of uncertainty of several aspects of the supply chain. But there is little attention in hunger relief operations. The reviewed literature shows that supply and demand traceability and visualization of food insecurity as the KPIs considered in the literature focused in the category of reduction of uncertainty (Table 2.4). Davis et al. (2013) executed the use of time series estimating techniques, moving average and exponential smoothing to predict the amount of food donated per description of food per type of donor in the food relief operations of Foodbank of central and eastern North Carolina. According to their data analysis, exponential smoothing method had provided better prediction results than the other established methods to predict the food donation. An empirical model was introduced by Phillips et al. (2013) to predict the total quantity of food obtained by a food bank in North Central Colorado. The model presented is a threshold model where in, Generalized Pareto distribution is applied, and the food donated by food providers was modeled based on the characteristics of the donors. This study mainly focused on recognizing the gap between demand and supply and ways to tackle the gap. Jian et al. (2013) explored different data mining techniques to study the trend in the donation and stochasticity in the donation using Markov Chain analysis. However, these methods and studies do not take into consideration the nutritional focus of food banking and the various important categories of food as per the nutritional guidelines involved in a food bank network and did not consider the purchased food aspects of food banks. Another study by Brock et al. (2015) studied the predictive modeling of donations arriving from supermarkets considering only the supermarket sales and implementing traditional and non-traditional forecasting methods. Recently, Nair et al. (2017a), Davis et al. (2016) and Nuameh (2016) evaluated several different approximation methods to estimate the daily availability of food based on a set category of foods and food providers and considering only the correlation between food types donated. In these studies, too however, there is no research done towards having a nutritional focus of food banking, nor is there an account of food purchasing data that should be included in the predictive study as it is one of the main ways of procuring food in the food bank supply chain. By better understanding the dynamics and function of the food assistance network in a given area, Alotaik et al. (2017) and Sucharitha and Lee (2019) developed clustering algorithms to better understand the potential causes of food insecurity in a given region. Table 2.5 provides a concise description of each analysis.

Table 2.5 Recent Background about predictive studies for Food Bank supply chains

Reference	**Method Implemented/ Chosen as best fit**
(Philips et al., 2013)	Generalized Pareto Distribution
(Davis et al., 2013)	Time Series Forecasting Techniques
(Delpish & Jiang, 2019)	Parametric Machine Learning Algorithms
(Brock & Davis, 2015)	Parametric Machine Learning Algorithms
(Nuameh, 2016)	Parametric Machine Learning Algorithms
(Davis et al., 2016)	Parametric Machine Learning Algorithms
(Nair et al., 2017)	Parametric Machine Learning Algorithms
(Alotaik et al., 2017)	Cluster Analysis
(Sucharitha & Lee, 2019)	Cluster Analysis

2.2.4 Food Quality, Nutrition and Reduction of Food Waste

An added effort taken by the Food Banks is to minimize food wastes which are often contradictory to the primary objective of Equity. Food Banks are infrequently faced with the interesting situation where they have some extra food on hand. This food may be produced with a lower shelf life and hence, has to be distributed in a competent manner. If a produce item's shelf life is ending, they face the conflict of whether they should send that food to waste (which also costs money) or send the extra food to an agency available. Faced with a situation like this, they choose to send food to the agency, although it may have a higher cost and violate equity. Food banks face challenging decisions like these continually, as mentioned by Orgut et al. (2016) (KPIs in Table 2.2). Gharehyakheh et al. (2019) consider the effect of temperature in a kinetic model to predict the remaining shelf-life of perishable foods in food banks. This method lessens food waste and improves food safety, thereby enabling food banks to serve more people in need. Sucharitha et al. (2018) proposed a simulation design and algorithm to ensure equitable and fair distribution of donated foods as the demand requirements and ensuring minimum wastage of food with a focus towards nutrition. Gharehyakheh (2018) developed an optimization model that maximizes food collection considering the type of food based on a recommended dietary guideline, subject to transportation and warehouse capacity constraints. To combat food wastage and ensure quality nutrition, Gonzalez-torr and Coque (2016) developed a Data Envelopment Analysis (DEA), in order to learn about food bank visitors and determine some clues about the efficiency of the food bank operations, comparing them according to variables such as the number of volunteers and permanent staff, the tonnage of food delivered, and the number of people served. A systems dynamics model was implemented by Galli et al. (2019) with the objective to set relations and dynamic mechanisms associated with variables relevant to food waste generation, food improvement for food relief operations. The analysis of feedback interactions highlights the (actual and potential) vulnerabilities of food assistance systems that occur when addressing food poverty by reducing food surplus and food wastage. To observe the contradictions with food wastage and Equity, Balcik et al. (2014) and Ortuño and Padilla (2017) developed mathematical models to minimize food wastage and maximization of equity as objective functions to study the results of a food bank case studies using heuristics.

2.2.5 Cost Reduction

Funding in terms of government and private means for the food banks and their respective food agencies in the Feeding America network can be obtained through multiple means. Some of their funding sources are local, state and federal governments and donations from individuals, religious institutions or companies. The food banks use the budget mainly to pay for their overhead expenses, acquire food, and ensure suitable transportation of donated food and wage for their staff. The economic condition also has a significant effect on their staffing levels and logistics capabilities. For these reasons, food banks must budget their costs and assets to ensure the longevity of their ability to obtain and distribute food to serve the food insecure population (KPIs for this category provided in Table 2.2.). The first introduction towards cost reduction related to a non-profit food relief organization such as food banks was presented by Solak et al. (2014), which provided a tactical decision system in the food bank distribution network to optimize delivery sites, agency assignments, and vehicle routing from these delivery sites with transportation costs as the minimizing objective function. A pickup and delivery model with backhauls was proposed by Davis et al. (2014), integrating food safety to schedule the routing for a food bank distribution network in the northwest region of North Carolina. Here, the cost reduction is related to the transportation constrictions. Several extensions of the routing models related to transportation costs have been developed (Aiello et al., 2015; Nair et al., 2016a; Rehaineh et al., 2017; Nair et al., 2016b). Schneider et al. (2019) developed a multi-criterion capacitated vehicle routing with multiple time windows approach to improve the auditing efficiency and developed multiple objective functions to minimize costs and maximize audit time. The model is evaluated using both exact and heuristic methods and analyzed the trade-offs between competing objectives. Ahire et al. (2018) implemented an integer programming optimization model to determine the optimal number of events of each initiative per year to maximize the total annual yield of meals (i.e., the number of meals that is provided using the cash donations and food rescued), subject to constraints of certain resources and the allowable number of events of each initiative. Glover et al. (2014) developed a continuous improvement project towards the food bank logistics cost reduction, ensuring order fulfilment and lead time reduction. Balza-Franco et al. (2017) developed a simulation model of the food bank logistic network and developed several scenarios using a game theory approach to minimize the logistics costs and find the best scenario and optimum logistics model.

Effective decision-making can be achieved through effective data visualization and visual analytics. The visual display and interpretation of data are vital to getting valuable intelligence that lies beyond quantitative analysis. Visualization tools depict intricate patterns that cannot be confirmed in any other form. Desai (2015) developed interactive dashboards to enable judicious decision-making for optimal food bank operations to meet hunger needs. Visual analytics was further extended by Hindle and Vidgen (2018), where an analytics methodology was provided to develop a dynamics visualization tool.

Gleaning and volunteering staff programs organize volunteers to ensure the food bank network's optimality to feed food-insecure individuals. Thus, the volunteering process simultaneously reduces food waste and food insecurity. However, this process's implementation and maintenance are challenging because volunteering relies on two uncertain input sources: food and labour supplies. Sönmez et al. (2016) first contributed to the emerging literature by developing a stochastic optimization model that characterizes a food bank gleaning operation. The paper contributes to the literature on food bank supply chain operations by explicitly modelling the gleaning process (i.e., the stochastic arrival of gleaning opportunities and gleaner attendance). Lee et al. (2017) extended this model to determine the schedule that maximizes the volume of excess food rescued from food sources by modelling the gleaning as an operation of service, where donations arrive randomly hence, a need to be scheduled within a limited time window. The feature that differentiates gleaning operations from other service settings is that there is uncertainty in both when donations will arrive and the gleaners' attendance who are not obliged to attend gleaning trips. The model optimizes the gleaning schedule to maximize the expected total volume gleaned and determines under which conditions different operational strategies can help improve the gleaning operation's performance. Ata et al. (2019) developed a dynamic volunteer-staffing policy that maximizes the payoff associated with the amount of food gleaned. The suggested optimal policy is a nested threshold policy that specifies each donation class's staffing level (i.e., a donation of a particular food source type and donation quantity).

Table 2.6 Review of Cost reduction models

Method/Model	References	Summary
MIP Model	(Solak et al., 2012)	Provided a tactical decision system in the food bank distribution network with the aim of optimizing delivery sites, agency assignments and vehicle routing from these delivery sites with transportation costs as the minimizing objective function
MIP Model	(Davis et al., 2014)	Developed transportation schedules for collection of food donations and delivery of food-to-food agencies and introduced a set covering model to determine the assignment of agencies to an FDP, identifying optimal assignments
MIP Model	(Aiello et al., 2017)	Mathematical model for the coordination of the supply chain operating a food recovery policy
MIP Model	(Nair et al., 2016)	Periodic unpaired pickup and delivery problem was developed with the objective of transportation cost minimization, and a suitable heuristic algorithm was developed to solve the problem
MIP Model	(Reihaneh et al., 2017)	Developed an efficient multi-start heuristic that iteratively constructs the initial solutions to the vehicle routing and allocation problem, thereby outperforming the alternative optimization-based heuristics
MIP Model	(Nair et al., 2018)	Routing model that aims at simultaneously selecting a combination for each food provider and welfare agency and routes to meet their required service levels minimizing the total transportation cost
MOLP Model	(Schneider et al., 2019)	Multi-criteria capacitated vehicle routing with multiple time windows approach to improve the efficiency of the auditing schedule and developed multiple objective functions to minimize costs and maximizing audit time
MIP Model	(Ahire et al., 2017)	Created an integer programming optimization model to measure the optimal events of every initiative in a year to maximize the total annual yield of meals
Analytical Model	(Glover et al., 2014)	Continuous improvement project towards the food bank logistics cost reduction ensuring order fulfillment and lead time reduction
Simulation Model	(Balza-Franco et al., 2017)	Simulation model of the food bank logistic network and developed several scenarios using game theory approach to minimize the logistics costs and find the best scenario

Table 2.6 continued

Analytical Model	(Desai, 2015)	Interactive dashboards to enable judicious decision making for optimal food bank operations to meet hunger needs
Analytical Model	(Hindle & Vidgen, 2018)	Interactive dashboards to provide a dynamics visualization tool
Stochastic Programming	(Sönmez et al., 2015)	Stochastic optimization model that characterizes a food bank gleaning operation
Simulation Model	(Lee et al., 2017)	Stochastic optimization model to measure the schedule that maximizes the volume of excess food rescued from food sources by modelling the gleaning as a service operation where donation orders appear randomly soliciting to be organized within a limited time window
Simulation Model	(Ata et al., 2019)	Dynamic volunteer-staffing policy that maximizes the payoff associated with the amount of food gleaned

MILP: Mixed Integer Programming; MOLP: Multi-objective Linear Programming

2.3 Discussion and Research Gap Analysis

Research gaps were identified and evaluated based on the literature review on categorized issues for the report. The following sub-sections address a summary of the results as well as study discrepancies.

2.3.1 Improved fairness and sustainability

Food bank supply chains has the challenge to balance impacts of effectiveness, efficiency and equity. While there is prior work addressing the trade-offs between these objectives, there are still opportunities to develop operations research and systems engineering techniques to enhance the feasibility of implementing the results to practical fruition by the food bank officials. Providing guidelines and training to the food bank officials on the handling of these trade-off weights and decision making should be considered. Furthermore, previous research was focused on development of deterministic and stochastic models to address equity objectives on a single time period as shown in Section 2.2.3. Whereas food banks generally have to plan over a longer time period. Another drawback in previous research is the lack of consideration of multiple food products to observe the equity and other cost reduction objectives and their trade-offs and also to observe the correlation and interactions between varying food groups or items.

2.3.2 Food Quality, Nutrition and Reduction of Food Waste

The regional foodbanks collect the donated, purchased, and government-supplied foods, and distribute the foods to approximately 60,000 locally affiliated agencies that provide groceries and hot meals to low-income families through food donation programs, such as food pantries and soup kitchens. The core principles governing the distribution of foods to the agencies are “as much as possible” and “as equitable as possible.” Accordingly, the food distribution performance of a foodbank is typically evaluated in total pounds distributed and distribution equity among counties (or agencies) only despite the fact that the nutritional quality of the foods distributed by the foodbanks has the potential to make a critical difference to the health of recipients. This further leads to an unreasonable outcome from the aforementioned instance. Together with concerns about obesity and diet-related chronic diseases such as diabetes, food banks are becoming more conscious of the need to increase the dietary content of donated foods, in the form of nutrition profiling (quantitatively score the nutritional value), nutrition policies (guide efforts to eliminate unhealthy products such as soda or candy), or fresh produce (increase to fill the nutritional gaps). However, little attention has been paid on how to distribute foods in a nutrition-rich or nutrition-balanced manner as shown in Section 2.2.4. As shown in an observational study with 269 food pantries supplied from two large foodbanks in Minnesota in 2013, the nutrition quality (measured in HEI-2010 that will be discussed later) ranged from 28 to 82 out of 100 (Nanney et al., 2016). This large variability could be because of the lack of systematic consideration of nutrition quality in the current metrics of distribution performance. No comprehensive framework was found for decision making with respect to implementing food quality and nutrition along with the objective of food waste reduction. In addition to this, having multiple food commodities as decision variable in the mathematical models was not considered. Hence, it is suggested to consider various food types with their respective nutritional and quality aspects into mathematical and quantitative models as opportunity for future research.

2.3.3 Reduction of Uncertainty

According to the Salesforce Nonprofit Trends Survey, more than half of nonprofits (53 %) find it convenient to gather program data (including food aid services and other forms of nonprofits) (Salesforce.org, 2020). However, putting the information to use is more difficult. Fewer than half (47%) say analyzing the data is easy, resulting in a slew of challenges when it comes to monitoring and quantifying things like effect and efficiency. Furthermore, only 41% believe it is simple to use data to improve the overall effect of programs. Regardless of the fact that charities are getting more mobile with each passing year, only 29% claim they can quickly collect and view data via a mobile device. This highlights the relevance of analytics software for nonprofits, but according to the survey, only 45 % of nonprofits actually use analytics, with another 30 % aiming to do so over the next two years.

Nonprofits have vast databases that they can use to build mathematical models that will aid in fundraising optimization. As seen in Section 2.2.3, this concentration is lacking in the current research pattern. They will develop a more refined targeting approach by using segmentation and predictive analytics to define and target the right groups. Data analytics assist NGOs in identifying and categorizing donors based on a number of criteria, helping them to better focus their messaging and fundraising activities. Data collection also assists charities in identifying partnerships that can be used to establish particular incentives.

Nonprofits, such as food banks and food agencies, must be able to work during the year. They must be effective in their activities and donor outreach in order to do this. On those fronts, data science research will aid, and current research (as seen in Section 2.2.3) shows that charities need those solutions. Food assistance services can use analytics to boost the budgeting process, streamline processes, maximize cost savings, assess and refine financial margin by service, model and forecast results (e.g., membership patterns, donor trends, resource needs, and revenue expectations), model and forecast success (e.g., membership trends, donor trends, resource needs, and revenue expectations), and strengthen overall mission effectiveness. Nonprofits may benefit from such solutions in a variety of ways, including the potential to monitor return on investment. Nonprofits may use machine learning to determine the best opportunities for making a donation. To forecast possible donation habits of prospective donors, modeling can be used to analyze donor profiles, user metrics, and relationships with an institution. The opportunities with the most money to donate will then be identified using wealth sampling and publicly accessible records. These resources, when used together, will help the organization narrow the reach of the target database to identify opportunities for project personnel to deal with, make more knowledgeable contribution demands, and discover new prospects to fill the donor stream. Analytics can also be used to enhance a firm's membership recruiting and fulfillment procedures. Data processing and visualization can also help with real-time monitoring and recovery operations during disasters. These organizations will use data mining and machine learning to ensure that they're investing their time, commitment, resources, and energy in the right places. Nonprofit groups need funding to carry out their missions, but they must also demonstrate the impact of their efforts in order to draw donations. Data science holds tremendous promise for minimizing this loop, reducing uncertainty, and assisting organizations in making well-informed, rational decisions.

It is undeniable that data science has tremendous potential for both large and small government and nonprofit organizations. However, having access to experts is a popular challenge for these organizations. Processes may be cumbersome, and finances can be insufficient to hire a full-fledged staff. For organizations, one option is to employ freelance data science professionals for short-term assignments. This allows them to gain access to data analysts' skills while making the process cost-effective. This is one way the third sector will profit from data science while still remaining motivated so that they can empower others. Also, there has been scarce work done towards studying the ChoiceTM system developed to auction for food based on online shares. Developing analytical tools to access the best possible food resources in an uncertain supply situation is key towards ensuring continued supply of food to the food insecure.

2.3.4 Cost Reduction

Various operational and strategic decision making, and quantitative models have been developed for the reduction of costs and improvement of planning and scheduling with respect to the food bank supply chain (As shown in Section 2.2.5). Most of the mechanisms implemented for cost reduction, have been utilized or sourced from for-profit organizations as inspiration. However, food banks and hunger relief organizations present a unique perspective of handling high amounts of vagueness with respect to supply, staffing, and donations. Hence, the fundamental challenge of ensuring suitable workforce or resources (i.e., volunteers or vehicles) to ensure demand is equitably fulfilled should be considered by incorporating aspects of supply, demand, and capacity constraints into a single study and observe the impact on planning scenarios. Additionally, while there is prior work in addressing cost reduction planning and scheduling models in food bank distribution system, there are still opportunities to develop new policies to enhance the planning system in food banks. The current groundwork does not take into consideration the possibility of multiple warehouses in their routing strategies, multiple vehicle capacities and types, types of food products, and the need for stochastic scenarios to observe the dynamic environment which food banks indefinitely face on a daily basis.

2.4 Conclusion

In summary, there is a growing body of research addressing logistics and supply chain management issues in the area of food and hunger relief with the focus of involving engineering and optimization approaches to improve food accessibility, ensuring the equitable, efficient, and effective functioning of the food bank network. This review is focused on critical issues which are underrepresented in the past literature reviews. Through a systematic and structured literature review, the article provides insights into the conceptualization and modelling ideas on the issues related to fairness and sustainability, cost reduction, food quality and nutrition, and data uncertainty. In this perspective, a total 48 previous published research articles were selected, categorized, and reviewed to find the research gaps and future research scope. The research gaps were identified, discussed, and suggestions were made for future research opportunities. The opportunities for future research will help academicians, practitioners, and researchers in their future work. Classifications and cited references may be used as a broad frame of reference to develop concepts and models for future research. Our study's particular limitation is that only published journal and conference articles were considered for the review. This work may be further extended for the study of food bank logistics from a systems perspective.

3. FORECASTING DONATED FOOD SUPPLY IN FOOD BANK LOGISTICS USING A HYBRID METHODOLOGY: A DATA-DRIVEN ANALYSIS

3.1 Introduction

Humanitarian relief and the alleviation of the needy is the help given to reduce the anguish created by natural or man-made disasters to human lives. Amid such emergency cases, hunger and starvation are among the major demands that emerge as a challenge to the local authorities in each affected region. Along with these effects, socio-economic components and factors also contribute to the emergency crisis. To counteract hunger-based crisis, assistance is funded by donations from individuals, organizations, governments, and other international groups. In addition to this, chronic hunger is a silent emergency. Each day, families globally struggle to feed themselves and their children with essential nourishment. There are approximately 925 million people in the world facing hunger (Feeding America, 2015). Around 1.4 billion people live on less than $ 1.25 U. S dollars a day.

On the global scale, the United Nations has initiated the world's largest humanitarian organization addressing hunger called the "World Food Program" (WFP) to promote food security (Galli et al., 2019). *Food security* is defined as a contact to sufficient food for an active and healthy life for all household members throughout the year (USDA, 2020). The WFP supplies food to 90 million people per year, of whom 58 million are children. The WFP is a member of the United Nations Development Group and works through its offices in 80 countries worldwide to assist people with food shortages (Kaiser, 2011). The WFP aims to solve global hunger by raising support within the United States based on individual and corporate efforts to shape public policy to eliminate hunger (Orgut et al., 2016b). Charitable organizations are often pushed to identify the critical evidence regarding a crisis and answer the significant gaps to address the current situation. These challenges necessitate a holistic view of the situation, which may be achieved through data acquisition and knowledge management. Well-informed rapid responses by collecting, distilling, analyzing, and managing the vast amount of information will lead to effective planning and response to relief operations.

In the United States, an estimated 14 percent of households were food insecure in 2014 (Feeding America, 2015), and the number of food-insecure households is expected to increase due to the COVID-19 pandemic (Providence, 2020). Along with food insecurity, food wastage is also on the rise. In the United States, around 40 percent of food goes to waste and ends up in landfills, which leads to environmental degradation due to the vast methane emissions from the landfills (Wittman et al., 2017). This quantifies to 20 pounds of food per person per month in food waste. According to a recent study, reducing food wastage in the United States by 15 percent would provide enough food to feed the country's food insecure population (Galli et al., 2019). As a result, increasing the effectiveness of hunger relief requires a concerted effort from individuals, businesses, and the government. Feeding America is one of the largest and well-known hunger-relief organizations in the country working towards reducing hunger. This organization strives towards distributing food by working collaboratively with the local consumers and the local governments by obtaining food donations and raising funds (Feeding America, 2015). Over 200 food banks under the Feeding America network are serving counties around the country and are feeding over 46.5 million Americans, including 12 million children and seven million seniors every year. The availability of food donations is highly unpredictable due to the charitable nature of the food bank network and the donation-driven environment, making this donation-driven environment considerably different from commercial supply chains that are more profit-focused. While the food banks of Feeding America aim to maximize and optimize the hunger-relief and minimizing food waste, the movement of food distribution should be done cost-effectively and efficiently.

The intent of this paper is multifold. The foremost aim is to analyze a particular food bank supply chain with a unique emphasis on forecasting food donations' supply side inputs. Several studies in previous literature examine food bank usage and the tasks associated with limited and unpredictable supply (Campbell, 1991; Tarasuk et al., 2019) and implementing studies to study the nutritional condition of donated foods and their availability (Handforth et al., 2013). However, to the best of our comprehension and findings, the application of statistical analysis techniques to handle the time series data with a hybrid methodology and a nutritional focus of data handling has not been addressed. We fill this gap by explicitly studying the nature of food donations and the uncertainty of the supply of food donations and implement hybrid methodologies of time series forecasting techniques to handle the dynamic properties of the food supply variable being measured. In particular, Autoregressive integrated moving average (ARIMA), Support Vector Machines (SVM), and Neural Network Auto-Regressive (NNAR) are the approaches utilized. This paper employs time series models to capture the dynamic nature of the donation behavior in hunger-relief operations. The assessment of which method best forecasts the donation behavior is done by comparing the models' accuracy when doing walk-forward cross-validation.

Our study has specific merit because it is essential for non-profit organizations to leverage knowledge and technology to reinvent their operational effectiveness. Food banks equipped with better predictive information on supply donation behavior can refine their donation strategies and planning at various levels of their logistics (Strategic, Tactical, and Operational) and make conversant decisions, which in combination increases the potential of the supply chain to meet organizational objectives earlier mentioned. The remainder of this study is indicated as follows: Section 3.2 summarizes the related literature. Section 3.3 outlines our methodology to analyze and estimate the behavior of food donations. The results of the predictive models of donation behavior are discussed in Section 3.4. Section 3.5 and Section 3.6 provide some concluding remarks about the implication of our results on efficiency and service provided by the food bank network to the food insecure.

3.2 Literature Review

In different fields, there is a considerable amount of related literature discussing the role of forecasting techniques in predicting potential demand using historical evidence. Although the majority of them focus on domains such as transportation planning, financial-based forecasting, weather forecasting, and so on, just a few mention the use of forecasting and machine learning techniques in estimating blood donation demand and supply (Drackley 2010; Ferreira et al., 2018), scarce resource consumption (Holgun-Veras et al., 2012; Amorim et al., 2018), and potential organ donation (Schleich et al., 2013). Forecasting models have gained little attention in hunger-relief activities, despite their strong applicability. While a few recent studies have concentrated on fair allocation of donated food (Orgut et al., 2017; Orgut et al., 2016a) and, improving distribution and collection schedules (Reihaneh & Ghoniem, 2017; Eisenhandler & Tzur, 2019b; Davis et al., 2014; Rey et al., 2015.; Nair et al., 2016), few have discussed the need for predicting the food donations. Phillips et al. (2013) proposed an observational approach to estimate the average volume of food saved by the Colorado Food Bank. The authors used a peak over threshold model to characterize the food donation mechanism, with the events greater than zero modeled using a Generalized Pareto distribution. Food provider's surplus food was modeled as a result of their form of donation (grocer, manufacturer, individual and farm). Their studies concentrated on assessing the difference between demand and availability, as well as methods for improving the overall quality of donated food. However, only the overall amount of food saved was taken into account, with no focus on diet or the nutrient importance or type of food rescued. Lien et al. (2014) suggested a resource allocation model for a hunger-relief organization in Chicago, based on an egalitarian welfare utility function as an indicator of fairness, for efficient and balanced allocation of donated food. The developed allocation model was compared to instances of known and unknown food supply prior to routing, and the results revealed that only when the supply of food is known prior to routing was equity maximized and wastage minimized optimally. Davis et al. (2016) looked at food rescue operations depending on the type of food and the type of donors that provided the food. In comparison to the moving average process, the results showed that exponential smoothing provides a good forecast accuracy of food donation. Jiang et al. (2013) used Markov Chain analysis to investigate various data mining methods to research the pattern of donation, the effect of a donor's frequency of donation on the overall sum of donation, the trend in donation, and the stochasticity in donation. These models, however, were limited to assessing average monthly food supply. Brock and Davis (2015) expanded this analysis by using a number of machine learning algorithms to estimate the total daily donation. They contrasted the results of traditional forecasting approaches with a data mining technique using a multi-layer perceptron neural network (MLP-NN) to determine how much food the Food Bank of Central and Eastern North Carolina will collect from a supermarket. In their study, the key emphasis of prediction was on supermarket donations, and donations were thought to be a result of supermarket sales. These initiatives to predict potential food contributions and the origins of each contributed food commodity have broadened our awareness and exposed new difficulties in predicting food insecurity.

For forecasting linear time series, the Autoregressive integrated moving average (ARIMA) is a common choice among conventional models. For nonlinear data structures in time series data, machine learning models such as artificial neural networks (ANN), support vector machines (SVM), and Long Short-Term Memory (LSTM) have been shown to perform well. According to previous research and our results, food donation supply is neither strictly linear nor nonlinear. There are typically both linear and nonlinear variations in them. If this is the case, then the ARIMA or ANN as a whole is insufficient to model those scenarios. As a consequence, integrating linear and nonlinear models to effectively model such complex autocorrelation systems can be helpful. In previous literature, several hybrid methodologies were investigated to solve a variety of time series problems, especially in financial stocks, econometrics, electrical energy, and other applicable fields. (Pai & Lin, 2005; Khashei & Bijari, 2011; Zhu & Wei, 2013; Arora & Taylor, 2016; Kordanuli et al., 2017; Tümer & Akkuş, 2018.; Maleki et al., 2018; Chakraborty et al., 2019; Mehtab & Sen, n.d). The hybrid ARIMA-ANN model (Zhang, 2003) has gained popularity for its ability to reliably predict complex time series. The disadvantage of this hybrid ARIMA-ANN paradigm is that it needs arbitrary reasoning to choose the number of hidden layers in the ANN architecture. To overcome this limitation, we focused on creating models that are more ‘‘white-box-like" in nature. NNAR suits a feed-forward neural network model with just one hidden layer to a time series with lagged values of the given series as inputs, according to our assessment (also flexible to handle some other exogenous data). NNAR has the advantage over ANN, SVM, and LSTM in certain cases because it is a nonlinear autoregressive model with less uncertainty, simpler interpretability, and improved estimation. As a result, NNAR is gaining more coverage in recent non-stationary time series forecasting literature (Maleki et al., 2018). The key aim of this paper is to work with time series data and make some decisions on how to characterize the recent dynamics of the observed values of the time series. Taking a final policy decision based on a component model can be counterproductive to overall preparation and fair distribution of donated food to the food insecure, particularly in view of the COVID-19 pandemic, where shifts in the complex properties of the element being evaluated are frequently observed. The most popular approach to this issue is hybridization of two or more models, which takes advantage of model diversity to minimize both the bias and variances of the prediction error achieved using single models (Oliveira & Torgo, 2014).

When the full data characteristics are unknown, hybrid models are more accurate (“Combining Pattern Classifiers: Methods and Algorithms - Ludmila I. Kuncheva - Google Books”, n.d.). In this article, we create a hybrid ARIMA-SVM and hybrid ARIMA-NNAR model that captures dynamic data structures and linear and nonlinear behaviour of donated food supply data sets, as a result of these discussions. ARIMA captures the linear patterns of the data collection in the first iteration of our proposed model. Then, using residual values obtained from the base ARIMA model, the NNAR and SVM models are used to capture nonlinear trends in the data. The proposed model is simple to understand, has a high level of predictability, and can adjust seasonality indices. We have shown the excellent performance of the proposed hybrid models for food supply forecasting for a specific food bank agency and differing information classes of the food supply dataset by experimental evaluation (thirteen difference information classes hence thirteen different datasets).

3.3 Methodology

The drawbacks of single time series models can be overcome using a hybrid approach. Traditional time series forecasting methods require stationarity in both the mean and variance, but machine learning methods have the capability of effectively modeling any type of data pattern and can thus be applied to the original data (Gorr, 1994). A combination of linear and non-linear time series models is often used to highlight the salient features of data sets in the process of capturing typical trends in the data. In this literature, hybrid ARIMA-ANN (Khashei & Bijari, 2011) and hybrid ARIMA-SVM model (Pai & Lin, 2005) are two common hybrid models where the main motivation was to understand both linear and nonlinear patterns of the data. These models have demonstrated superior performance in terms of prediction accuracy for a variety of forecasting problems in economics, sales, finance, carbon price, stock, oil, and other fields. In this paper, we propose a novel ARIMA and NNAR walk-forward hybridization modeling to solve the problem of forecasting donated food supply. The component models used in the hybridization are described briefly below, as they were found to better capture the behavior of the food supply data.

3.3.1 ARIMA model

Box and Jenkin (2011) introduced the ARIMA model, which is a linear regression model used to track the linear tendencies in stationary time series data. ARIMA(p,d,q) is the model's name, with p, d, and q being integer parameter values that define the model's structure. Specifically, p and q denote the order of the AR and MA models, respectively, and d denotes the level of difference applied to the given data.

The ARIMA model explicitly models both the autocorrelations of the series values and the autocorrelations of the forecast error. The following is the ARIMA model's mathematical expression:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q},$$

where y_t denotes the actual value, ϵ_t denotes the random error at time t, ϕ_i and θ_j are the coefficients of the model. The independent and identical (i.i.d) condition is assumed to hold for ϵ_{t-1} with a zero mean and constant variance. The methodology consists of three main iterative steps: (1) model identification and model selection; (2) parameter estimation of the model parameters, (3) model diagnostics checking (namely, residual analysis) are performed to find the 'best' fitted model. Differentiation is used once or twice in the model identification and model selection steps to achieve stationarity for non-stationary data. When the stationarity condition is met, the autocorrelation function (ACF) plot and partial autocorrelation function (PACF) plot are used to choose between AR and MA model groups. The parameter estimation stage entails an optimization process that makes use of metrics like the Akaike Information Criterion (AIC) and/or the Bayesian Information Criterion (BIC). Finally, the residual analysis is performed in the model testing phase to evaluate the 'best' suited ARIMA model. The ARIMA model is a data-dependent approach that adapts to the data set's structure. However, any large nonlinear data set will constrain the ARIMA model, which is a significant disadvantage. As a consequence, for forecasting complex time series structures, the proposed hybrid model employs the NNAR model to deal with nonlinear data patterns.

3.3.2 NNAR model

Neural networks are complex nonlinear forecasting structures based on basic mathematical models of the brain. A neural network is a network of ‘‘neurons" organized in layers (viz. input, hidden and output layers). A linear combination of the inputs yields the forecasts. In the network model, the weights are chosen using a "learning algorithm" that minimizes the mean squared error. The NNAR model is a nonlinear time series model that uses lagged time series values as neural network inputs (“Forecasting: Principles and Practice - Rob J Hyndman, George Athanasopoulos - Google Books”, n.d.). The feed-forward neural network NNAR(p,k) has a single hidden layer with p lagged inputs and k nodes in the hidden layer. An NNAR(4,6) model, for example, is a neural network that forecasts the output with six neurons in the hidden layer implementing the last four observations $(y_{t-1}, y_{t-2}, \dots, y_{t-4})$ as inputs. The feed-forward neural network NNAR(p,k) has a single hidden layer with p lagged inputs and k nodes in the hidden layer. To ensure stationarity, this model can be applied to the original nonlinear data without any constraints on the parameters. For non-seasonal data sets, a NNAR(p,k) model uses p as the ideal number of lags for an AR(p) model, and k is set to $k = [\frac{p+1}{2}]$.

Formulation of the hybrid model

The ARIMA model is a well-known mathematical model for predicting linear time series. The NNAR model, on the other hand, can detect nonlinear patterns in the data collection. As a result, the two models are combined sequentially to account for both linear and nonlinear tendencies in the model (Chakraborty et al., 2019). For forecasting food supply and the various information classes used, a hybrid strategy with both linear and nonlinear modeling capabilities is a good choice. The ARIMA and NNAR models can capture data characteristics in linear and nonlinear domains in different ways. As a result, the hybrid method can model both linear and nonlinear trends, resulting in better overall forecasting results. There are several time series models in the literature, and many studies show that hybrid models increase forecast accuracy. The aim of designing a novel hybridization is to maximize the benefits of single models while lowering the probability of single model failure. The hybrid method based on linear and nonlinear model assumptions is based on the premise that the relationship between the linear and the nonlinear components is additive. Single model strength is crucial for hybridization, and this selection is necessary to demonstrate continuous progress over single models. This paper introduces a novel ARIMA-NNAR hybridization that overcomes the weaknesses of single models thus optimizing their strengths.

In contrast to component models, the proposed approach will guarantee better results by integrating linear and nonlinear models. The following is a representation of the hybrid model (Z_t):

$$Z_t = Y_t + N_t,$$

where Y_t is the linear part and $N_t$is the nonlinear part of the hybrid model. Both Y_t and N_t are estimated from the data set. Let, Y_t be the forecast value of the ARIMA model at time t and ϵ_t represent the residual at time t as obtained from the ARIMA model; then,

$$\epsilon_t = Z_t - \widehat{Y_t}\,,$$

The residuals are modeled by the NNAR model and can be represented as follows,

$$\epsilon_t = f(\,\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-n}) + \delta_t,$$

where f is a nonlinear function modeled by the NNAR approach and δ_t is the random error. Therefore, the combined forecast is,

$$\widehat{Z_t} = \widehat{Y_t} + \widehat{N_t},$$

where, $\widehat{N_t}$ is the forecast value of the NNAR model. The use of residuals in the diagnosis of the sufficiency of the proposed hybrid model is justified because the residuals still contain autocorrelation that ARIMA could not model. The NNAR model, which can capture the nonlinear autocorrelation relationship, is responsible for this function. In conclusion, the proposed hybrid ARIMA-NNAR model operates in two stages. The linear component of the model is analyzed using an ARIMA model in the first step. The residuals of the ARIMA model are then modeled using a NNAR model in the next step. The hybrid model also reduces model uncertainty in inferential statistics and time series forecasting. A flowchart of the hybrid ARIMA-NNAR model is presented in Figure 3.1.

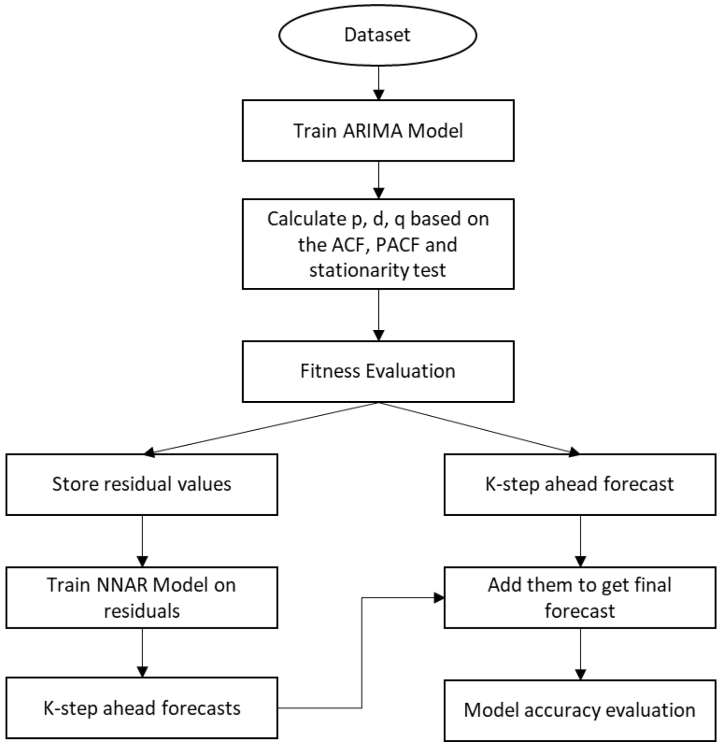


Figure 3.1 Flowchart of the proposed hybrid model

Although the model implementation is shown in Fig 3.1, we add an iterative cross-validation approach called walk-forward cross validation, also known as expanding window cross validation, to the hybrid technique, in which the parameters for each model can be varied and checked again on an iterative basis to obtain optimized tuning parameters unique to the algorithm.

The aim of time series forecasting is to predict accurate values ahead of time. Traditional cross validation methods used in machine learning, such as train-test splits and k-fold cross validation, do not function for time series results. This is because such approaches neglect the data's temporal characteristics (Phadke et al., 2020). Walk-forward cross validation technique is used to forecast at any time point, specifically an expanding window, to handle our current datasets efficiently and based on suitable time series predictor evaluation methods given in the literature (Schnaubelt, 2019). A model makes a prediction for the next time stage using training data in this technique. After that, the forecast is compared to the real value. The prediction's actual value is then applied to the training data, and the process is repeated by predicting the value for the next fixed time stage. For the final analysis, the average of each output metric considered for our study is calculated. For

N iterations of the entire time series, several train-test series are evaluated. The data was also tested using a sliding window walk-forward cross validation system, and the findings were compared to those obtained using an expanding window walk-forward cross validation method. The specifics can be found in the Appendix.

3.4 Data

A food bank organization provided three fiscal years of monthly supply data ranging from 2014 to 2017 at the start of our study. A fiscal year runs from July until June of the subsequent year. The data contains monthly food receipts by the donor. Each record contains several fields that describe specific information about the inventory transaction. However, we limit our data description to the key fields relevant to this study. The relevant fields are described in Table 3.1. The quantity (in pounds) of food received per receipt transaction is captured in the gross weight field and represents the dependent variable of interest for the forecast models. Each receipt transaction captures the specific source of the donation via the donor id field. Within this network, there are 1056 distinct donors. However, other descriptive measures about the source and type of donation would be useful for constructing forecasts. The donor type field serves to differentiate the donor by industry (e.g., retailers, restaurants, manufacturer). Each donor maps to a distinct donor type. Further information about the donated product itself can be identified by its storage requirements (storage type). The specific food type is classified following the MyPlate guidelines (MyPlate, n.d.) to ensure a nutrition-focused food banking decision-making in our forecasting analysis.

To study the relationship between the donation information and forecast accuracy, we develop thirteen information classes based on the relevant fields from the whole dataset (provided in Table 3.1). The forecasts are constructed for each of the information classes (shown in Appendix). The classes represent specific donation characteristics or fields for analysis. For example, a new time series constructed of food donations provided by manufacturers only can be extracted. This constitutes one information class. Each information class time-series dataset is formally defined in Table 3.2. We also consider an aggregated dataset with all the donation features in one dataset (C_N). The information classes developed were chosen based on the food bank management requirements mentioned in previous literature and based on ensuring the forecasting of the donated food supply with a nutritional view. This disaggregation of the data allows for analysis and prediction by various levels and fields. By providing analysis of various information class time series, food bank officials can observe the supply and ensure an equitable distribution of donated foods, especially since the food supply is scarce compared to the demand.

Table 3.1 Summary of key fields

Field	Description
Posting Date	The date the item was received
Donor ID	Unique identifier of the donor
Quantity Received	The amount received in pounds
Donor Type	Classification of the donor (e.g., manufacturer, retailer)
Storage Type	Type of storage of food
Food Type	Classification of donated food

3.5 Performance measures

Root mean square error (RMSE), Mean absolute error (MAE), and Mean absolute percentage error (MAPE) are the metrics used in this analysis to measure the efficiency of various forecasting models (including the proposed model) (MAPE).

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i\ |$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}\ |$$

where y_i is the target output, $\hat{y}_i$ is the prediction and n denote the number of data points. By convention, the lower the value of these metrics, the better the forecast model is. MAE gives the magnitude of the average absolute error. MAPE gives the percentage score of how forecasts deviate from actual values. It is useful for comparing performance across series of data that have different scales. RMSE is a widely used performance metric as it heavily penalizes bad forecasting. This measure has the same unit as that of the data series.

3.6 Analysis of results

The thirteen information classes as shown in Table 3.2 are provided with their respective time series datasets. We have studied ARIMA, SVM, NNAR model for this data. Each training data set has 36 observations from June 2014 to July 2017 and walk-forward cross-validation is implemented to assess the forecasting performance of the proposed model. The average of each performance metric taken to consideration for our study is measured for the final analysis. Multiple train-test series are evaluated for N iterations of the entire time series. We provide the time series plots of aggregated dataset along with the ACF and PACF plots are provided in Figure 3.3 and Figure 3.4. The remaining information classes data descriptions are provided Appendix.

Table 3.2 Information classes

Information Class	**Description**
C_N	All donations (aggregated)
C_{DM}	Donor type (Manufacturers)
C_{DR}	Donor type (Retailers)
C_{DRT}	Donor type (Restaurants
C_{DO}	Donor type (Others)
C_{SD}	Storage type (Dry)
C_{SF}	Storage type (Fresh)
C_{SFZ}	Storage type (Frozen)
$\_C_{FF}$	Food type (Fruits and vegetables)
C_{FM}	Food type (Meats)
C_{FB}	Food type (Breads and Cereals)
C_{FC}	Food type (Confectionaries)
C_{FD}	Food type (Dairy products)

We have applied our proposed hybrid ARIMA-NNAR model and other single and hybrid models to all the thirteen data sets as follows. Linear modeling is done with ARIMA(p,d,q) using the ‘‘forecast’’ package in R statistical software and stationarity testing and fitness evaluation. Nonlinear modeling with the NNAR approach is done with the ‘‘caret’’ package using the ‘‘nnetar’’ function and SVM with the Radial kernel function with the ‘‘e1071 package in R statistical software time series forecasting. Before fitting an ARIMA model, the order of the model must be specified. The ACF plot and the PACF plot aid the decision process. We then choose the ‘best’ fitted ARIMA model for each train data set. After fitting the ARIMA model, we generate predictions for every two months’ time steps to compute the residual value.

Further, ARIMA residuals are modeled with NNAR(p,k) model having a pre-defined Box-Cox transformation set to $\lambda = 0$ to ensure the forecast values stay positive. The value of p and k is obtained by training the network, which is indeed a data-dependent approach. Further, we add both the linear and nonlinear forecasts to obtain the final forecast results. The parameters for ARIMA, SVM, and NNAR are mentioned for each information class dataset. The aggregated information class details are provided in Table 3.4, and the remaining information classes are provided in the Appendix. Moreover, finally, we computed average test RMSE, MAE, MAPE after the iterative cross-validation technique of Walk-Forward and reported them. All the experimental results are reported in Appendix. The estimated values of the proposed model for the thirteen datasets and the actual test values with a 6-months ahead forecast on the food bank's request under study are depicted in the Appendix with the aggregated information class dataset results shown in this Chapter. From observing the forecasts, we see that the forecasts of the best models for each dataset appear to be good at predicting the general direction of the food supply.

The study and analysis reveal a few exciting time series characteristics in each of the data sets. As reported in this Chapter and the Appendix, the performance measures also reflect an inconsistency in forecast results. NNAR model shows the best result for some datasets, and our proposed hybrid ARIMA-NNAR model outperforms all other models for most of the datasets studied. The proposed model's performance is impressive among hybrid models, whereas hybrid ARIMA-SVM models seem not to perform well since they were mostly used for stock market forecasting and large data sets. This gives a guide to time series practitioners to understand the use of hybrid models. Overall, we can conclude from the analysis of the results that the proposed hybrid ARIMA-NNAR model either outperforms single and hybrid models or remains competitive for all the food supply information class data sets.

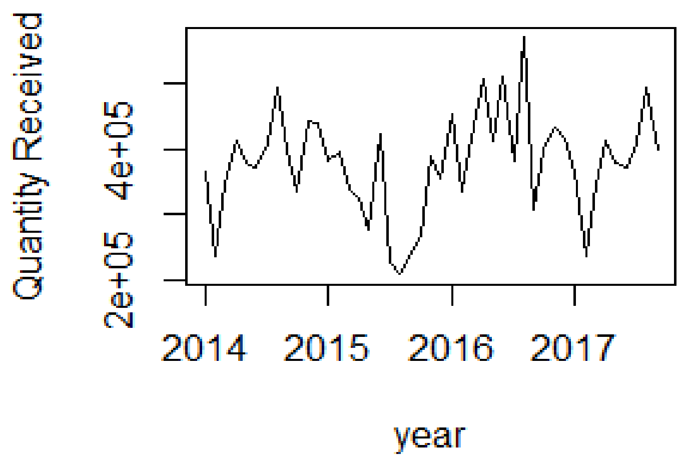


Figure 3.2 Training dataset for aggregated information class

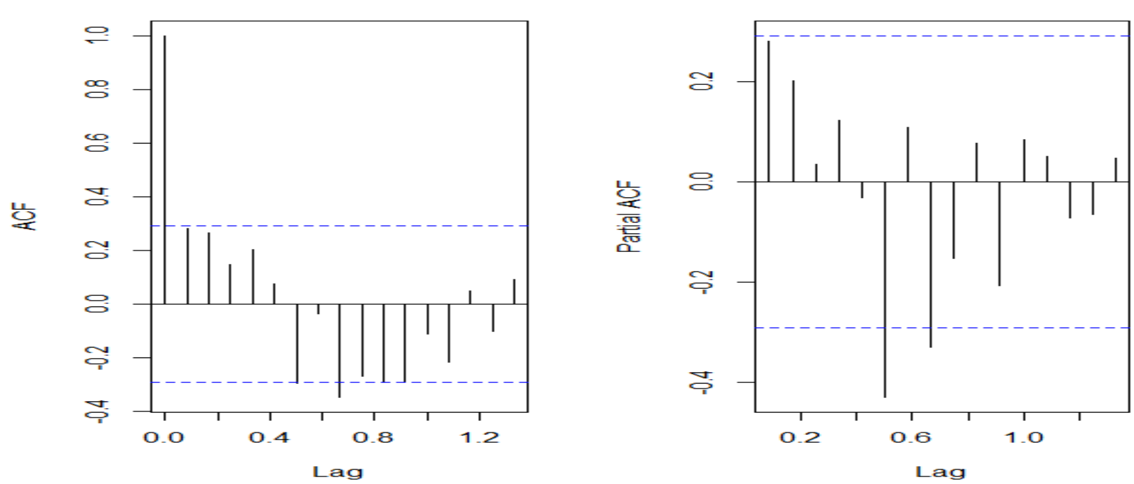


Figure 3.3 ACF, PACF plots for aggregated information class

Table 3.3 Quantitative measures of performance for different forecasting models for aggregated dataset

	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,1)	74972.76	57667.44	17.02
SVM	Gamma = 0.03, cost = 1	64169.81	52347.15	15.20
NNAR	(9,1,6)[12]	60648.12	47307.73	13.81
Hybrid ARIMA-SVM		39833.41	31266.62	9.32
Hybrid ARIMA-NNAR		**32358.59**	**28109.1**	**8.42**

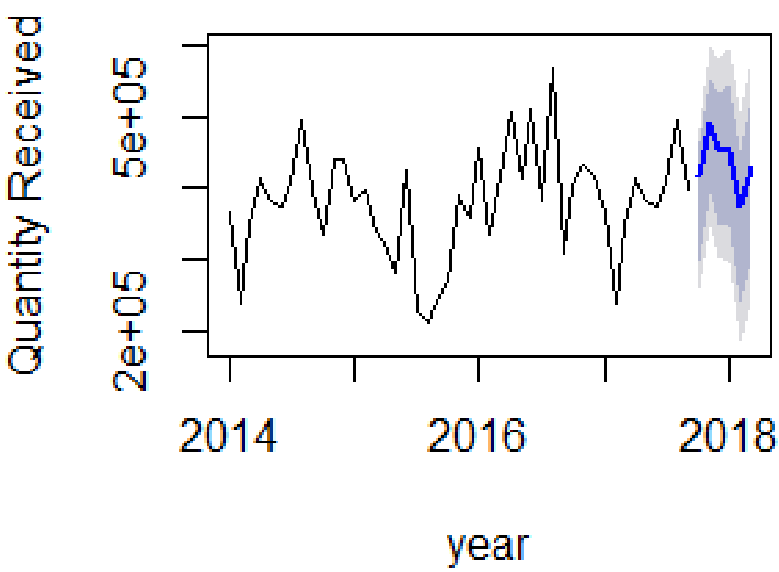


Figure 3.4 Actual vs predicted forecasts (using ARIMA-NNAR model) of the aggregated data set

3.7 Discussion

Depending on whether the time series is discrete or continuous, deterministic or stochastic, stationary or nonstationary, and linear or nonlinear, it can be classified. Trend, seasonality, stationarity, outlier, and residual analysis are the first steps in time series analysis, which are accompanied by the creation of a forecasting model based on the data set's characteristics (Divina et al., 2019). In practice, determining whether a time series under investigation is created by a linear or nonlinear underlying process can be difficult. Since all real-world time series data sets are complex in nature and often include both linear and nonlinear patterns, a single model may be inadequate to accurately address all of the data characteristics. As a result, a hybrid model that incorporates both linear and nonlinear components is beneficial. The additive relationship between the linear and nonlinear components of the time series is the fundamental assumption in the hybrid methodology (Domingos et al., 2019). Hybrid schemes work best for both stationary and nonstationary time series, as well as data with both linear and nonlinear patterns. We need the component model to be sub-optimal in order to improve the hybrid technique, and it would be useful to combine individual forecasts based on different information sets to generate superior forecasts. Our model is ideally adapted in circumstances where the data sets have enough nonlinearity and non-stationarity, based on our experience forecasting donated food supplies. Furthermore, our results show that expanding window cross-validation achieves better performance in terms of performance measures chosen than the rolling window cross-validation. Expanding window cross-validation also provides results without large computational costs as compared to rolling window cross-validation (Schnaubelt, 2019).

3.8 Conclusions

In this paper, we have built a hybrid model that performs superior for forecasting donated food supply. Our proposed hybrid ARIMA-NNAR model with the implementation of Walk-forward cross validation filters out linearity using the ARIMA model and predicts nonlinear tendencies with the NNAR approach. The hybrid ARIMA-NNAR model not only better describes the data's linear and nonlinear autocorrelation structures than conventional component and other hybrid models, but it also produces better forecast accuracy. The presumption of an additive relationship between linear and nonlinear components, however, is a drawback of the proposed methodology.

It is often true that no paradigm can be used uniformly in all situations, which is relevant to the ‘‘no free lunch theorem’’ (Bishop, 2006). Finally, we can conclude that our proposed model can help food bank policymakers to predict the subsequent planning methods for providing equitable distribution of donated food to the food insecure and respond to the donations and demand more effectively. Thus, this will reduce the uncertainty of food donations and will govern the employment of resources. Behavior of the proposed model for seasonal and multivariate time series datasets can be considered as a future research work of this paper.

4. DATA-DRIVEN FRAMEWORK OF FOOD DONATION AND PURCHASING BEHAVIOR FOR RESILIENT FOOD SECURITY

4.1 Introduction

In many developed nations such as United States of America, there are sections of the population that are more vulnerable to food insecurity even though these countries produce enough food to feed their people (Feeding America, 2014). Food security is characterized as a situation in which every person has physical, social, and economic access to sufficient, edible, and nutritious food that satisfies their nutrient needs and food preferences for a healthy and active lifestyle (Riches, 2002). The United States Department of Agriculture (USDA) estimates that food insecurity affects close to 15% of all U.S households and current reports have links with food insecurity and economic conditions, by this means, showing that food insecurity increased to 14.6% in 2008, with the beginning of the recession (USDA, 2020). Recently, several studies including the COVID impact survey, found that food insecurity has gone up drastically since the pandemic began (Providence, 2020). Every day, millions of Americans do not know when or where they will get their next meal which would hence lead them to be termed as food insecure.

The unpredictability and size of the issue of food insecurity does require several ways of mediations to viably mitigate its related hunger and health issues. Food insecure individuals receive assistance from the government, public and private organizations. Our focus is on one of the largest national non-profit hunger relief organization tackling hunger and food insecurity in the country namely, Feeding America. Feeding America offers food and aid to the underprivileged by setting up a nation-wide network of around 200 food banks and around 60,000 food pantries and meal programs (also termed as food agencies in the food bank supply chain). The set-up food banks obtain donated food from food donors such as retailers, manufacturers, wholesalers, restaurants, etc. The food being obtained from these donors undergo quality inspection before being distributed to the food agencies with the help of trucks and vans daily and periodically due to certain perishable food products collected that cannot be stored in food banks for a long period of time. It is from these food agencies where the food insecure people can receive donated foods (Bazerghi et al., 2016).

In addition to receiving donations in terms of food and other products, food banks also purchase foods for distribution to their food agencies. The monetary funds for purchasing foods are obtained through private monetary donations which is unpredictable at times, but it is through this donation, the food banks use to purchase items that their customers prefer or demand and also which the organization prefers due to the need for a better shelf-life and price (Bucknum & Bentzel, 2019). Predicting the behavior of purchasing of foods would aid in the better planning of food banks in ensuring storage capacity usage and efficiency for the future weeks of predicted purchasing of food from different sources and food types. Another reason why donated food and purchased food are dealt separately in our research is because, items purchased from monetary funds are done distinct and separately from items obtained from food donors (Orgut et al., 2016b). Another benefit of predicting the purchase behavior and focusing on purchased foods prediction separately, is to provide the respective food agencies associated with the food bank with the upcoming purchasing of food. Orgut et al. (2016a) also proposed that this is particularly helpful as purchased goods are made available to the food agencies for a small percentage markup over the procurement cost. The predictive results would prepare the agencies for working on their available funds to take advantage of the purchased food items. While food is purchased based on monetary funds, the type of food purchased is based on the demand of the food type which in turn leads to the usage of storage space of food that is consumed and not wasted due to lack of personal preference by the people in need (Chapnick et al., 2019). Efficiency and effectiveness of these hunger-relief organizations depends highly on the effective utilization and deployment of food with least wastage. A significant rationale for wastage and inequitable distribution of these foods is that the amount and class of hunger relief foods is obscure until it is seen by the truck/van driver's arrival (Nair et al., 2017a). Hence, the focus and aim of this study is to analyze and dissect the supply of donated foods and purchased foods to help non-profit organizations such as Feeding America to deal with this uncertainty. To tackle this ambiguity, it is imperative to see how uncertainty influences the logistics and activity of hunger-relief operations. Being a non-profit organization, profitability and cost efficiency is not the main or even, one of the important goals. Equitability, effectiveness, and fairness are of prime concern. Additionally, having a nutrition-focused food banking does aid in the reduction of food insecurity according to recent research (Wetherill et al., 2019). Hence, following the guidelines of United States Department of Agriculture and ensuring that the food products obtained, can be classified and categorized based on a particular nutrition-focused goal, will aid the decision makers in the organization to evaluate if they are ensuring or developing food security in their region allotted. There are different types of food products which are of interest to the food banks owing to the nutritional content, demand, and quantity of these food products. In general practice, the lack of donated and available private monetary funds for food supply information, leads to issues and delays in delivery of these rescued foods to the agencies which could lead to wastage of food as several food agencies have limited time slots of availability thereby leading to higher operating costs, food wastage, and unfair allocation of these foods (Bucknum & Bentzel, 2019). Hence, addressing these concerns are the primary objectives of numerous hunger-relief organizations.

Our study has several intents and contributions. The primary aim is to analyze datasets of food supply, one consisting of donated food products information and the other dataset consisting of purchased food products. The two datasets have been collected for the period of 2014-2017 from a specific food bank organization to obtain forecasting results of their supply of donated foods and purchased foods. Also, implementing several predictive models using various statistical learning methods which, to the best of our knowledge, the techniques, and applications of statistical analysis which we implement to study supply uncertainty has not been addressed. These models will be tested for predictive accuracy in order to choose the best model based on generalizability and ability to capture the data structure. Understanding the structure and pattern of available food supply, both donated and purchased, is crucial for successful planning and equitable food distribution among food agencies as sources of food for each dataset is different and the intent with which the food is obtained is completely different (Orgut et al., 2016a). It is our goal that the predictive models developed would prepare the decision makers in understanding the food availability better thereby also helping reduce operation costs. Hence the contributions of the study are as follows –

1. Development and predictive modeling of state-of-the-art statistical learning methods (parametric, semi-parametric and non-parametric models in literature)
2. Implementation of data sorting and classification based on the MyPlate guidelines to ensure nutrition-focused predictive results
3. Provide predictive information for efficient scheduling and planning based on the supply information of both donated and purchased food items

4. Introduction of a predictive framework that can be easily implemented by any food bank organization
5. Implementation of the predictive framework on case studies based on different levels of planning

The remainder of the article is structured as follows. Section 2 illustrates the background and a brief literature review. The datasets used in this study are then explained, and the explanatory variables are discussed in Section 3. Section 4 presents the methodology and brief description of different forecasting methods. Experimental results from different models are examined and compared in section 5 and discussions and case studies are provided in Section 6. Conclusions and future research directions are discussed in the final section.

4.2 Literature Review

Food bank supply chain falls under non-profit supply chains. Non-profit supply chains in general, show some important similarities towards humanitarian logistics. A more detailed examination of the objectives of non-profit supply chains such as food banks tell us a different mindset that explains humanitarian logistics and non-profit logistics. It is primarily for sudden onset disasters in humanitarian supply chains, where assistance is given for disaster-relief activities such as provision of food, water, and shelter (Balcik et al., 2008). (i.e., natural disasters like earthquakes, etc.). The aim of disaster relief organizations is to provide short-term, emergency assistance. In case of slow onset disasters like poverty, famines, etc., social-aid chains are present to capture supply chain activities that support social programs targeted for slow onset disasters such as homelessness, unemployment, poverty and crime. Clearly, social welfare and humanitarian relief chains have the same core objective of satisfying the needs of people affected by disasters (Balcik et al., 2008). However, according to Adivar et al. (2010), the two chains differ based on the occurrence that caused the humanitarian crisis and the length of aid. More specifically, social -aid chains embrace the goal of not just cutback but also the prevention of a social issue and operate continuously through social development and improvement. These organizations predominantly rely on monetary contributions and government assistance. Food bank supply chains fit into the humanitarian supply chain literature since they counter both imminent and delayed disasters such as job loss, pandemic scenarios, and hunger, as well as the tragedy of food insecurity. Like the objectives of humanitarian and social-aid supply chains, food bank supply chain's focus is not on profit, but on the importance of funding, donations, and equitable distributions.

With the recent exponential development in data gathered, as well as enhanced computational and data-processing capabilities, there is an increasing opportunity to use sophisticated statistical and machine learning algorithms to develop data-driven decision support models that can help inform policy to enhance overall sustainability tools (Wongso et al., 2019). More precisely, evaluating the varying supply of donated food and other goods and assisting decision makers in identifying the drivers of supply of various resources in the food bank sector and assisting decision makers in identifying the best areas of development and identification (in terms of productivity and conservation) for optimum return (in terms of socio-economic benefits) (Orgut et al., 2016b). There is significant pertinent literature examining the task of forecasting methods in estimating imminent demand using historical data in various fields. While most of them focus on areas like supply chain management, weather forecasting, sales, economic forecasting, etc., very little discussion is done in the study and the use of forecasting techniques in predicting blood donation demand and supply (Drackley, 2010; Drackley et al., 2012), possible organ donation (Schleich et al., 2013) and sparse resource utilization (Firat et al., 2009; Jain & Ormsbee, 2002; Nasr et al., 2002). Despite its wide applicability, forecasting models received little attention in hunger-relief operations. While the recent few studies focus on optimizing gathering and delivery schedules (Balcik et al., 2014; Fianu & Davis, 2018; Nair et al., 2016a; Nair et al., 2016b; Nair et al., 2017b; Solak et al., 2014) and equitable and effective allocation of rescued food (Balcik et al., 2014), few findings addressed the need of forecasting the donation of products to the food bank network. Research proposing that equity is maximized, and wastage is minimized when supply is known prior to routing has been studied and analyzed by Lien et al. (2014). In this paper, a resource allocation model is developed for a food rescue organization in Chicago, for effective and equitable allocation of rescued food, considering an egalitarian welfare utility function as an indicator of equity.

Davis et al. (2016) implemented the use of time series forecasting techniques, moving average and exponential smoothing to predict the amount of food donated per description of food per type of donor in the food rescue operations of Food Bank of central and eastern North Carolina. According to their data study, exponential smoothing method had provided better prediction results than the other tested methods to predict the food donation. Phillips et al. (2013) developed an analytical model to determine how much food a food bank in north central Colorado will get. The model presented is a threshold model in which Generalized Pareto distribution is used and food donated by food suppliers is modeled using the donors' characteristics. The main focus of this research was on defining the demand-supply gap and potential solutions. Using Markov Chain analysis, Jiang et al. (2013) investigated various data mining techniques to investigate the donation pattern and stochasticity. However, these approaches and studies ignore the nutritional emphasis of food banking and the various essential categories of food as described by nutritional guidelines in a food bank network, as well as purchased foods and those aspects of food banks where monetary contributions play a significant role in a non-profit organization (Malthouse, 2010). Brock and Davis (2015) looked at the statistical modeling of donations received from supermarkets using conventional and non-traditional forecasting approaches and only considering supermarket sales. Nair et al. (2017a) recently compared three separate approximation approaches for estimating daily food availability limited to a fixed category of foods and food providers, with only the association between food types donated being taken into account. However, there is no analysis done in these studies to include a nutritional emphasis on food banking, nor is there an account of food procurement data that should be included in the predictive study since it is one of the keyways of procuring food in a non-profit organization like food banks (Bucknum & Bentzel, 2019). Table 4.1 provides a brief description of each analysis.

Hence, in our study, we will implement and evaluate the latest parametric, semi-parametric and non-parametric statistical and machine learning algorithms on their ability to estimate the daily availability of food categorized as per the government approved nutrition guidelines of MyPlate some of which are donated by several donors and some are purchased by the food bank decision makers. We aim to bridge the gap of the literature by proposing a generalized probabilistic predictive framework- grounded in statistical learning theory to (a) develop an accurate predictive model, based on both in-sample-fit and out-of-sample predictive accuracy, (b) identify the important predictors of the food supply behavior. While a local food bank in Lafayette, Indiana is selected as a case study to demonstrate the applicability of the proposed framework, the methodologies presented in this paper are generalizable to other food banks and regions.

Table 4.1 Recent Background about predictive studies for Food Bank supply chains

Reference	**Method Implemented/ Chosen as best fit**
(Philips et al., 2013)	Generalized Pareto Distribution
(Davis et al., 2013)	Time Series Forecasting Techniques
(Jiang et al., 2013)	Cluster Analysis
(Brock & Davis, 2015)	Parametric Machine Learning Algorithms
(Nair et al, 2017)	Parametric Machine Learning Algorithms

4.3 Data

4.3.1 Data Description

In this section, we describe the data used to train our statistical models. The local food bank located in Lafayette, Indiana provided historical data of food supply. The datasets provided are of two types. One with the information regarding the donated food products and the other consisting of information regarding all the purchased food goods. Hence, in this research, we develop separate models, each for the donated foods dataset and for the purchased food items dataset. This way, the decision makers will be able to estimate the inventory and anticipation of food items based on donations and items needed for purchasing based on both the final predictive model's information. The data includes food supply information received from October 2014 to September 2017. The data consists of a total of 90,000+ records. Each record indicates the time stamp values, source of supply, food description, and quantity received. Considering that we are implementing machine learning models that are parametric, semi-parametric and non-parametric in nature, the possibility of trend and cyclic pattern of the datasets is left intact due to the multivariate inputs and varying factors influencing the response variable (Taneja, 2020). The input variables and the response variables are discussed in the below sections and listed in Table 4.2 and Table 4.3.

4.3.2 Donor data

The name of the donor and the name of the supermarket, manufacturer, or other entity that provided food were included in the donated food dataset. As a result, the data was converted to include the donors based on a specific identification number, as well as the categories of donors from which the food bank collects food.

4.3.3 Source data

The name of the source and the name of the supermarket, manufacturer, and other such place where the food was purchased were included in the purchased food dataset. As a result, the data was converted to include the sources based on a specific identifier as well as the type of buying source from which the food bank receives food..

4.3.4 Storage data

The datasets consisted of the name of the food item that was provided. This data was hence used to obtain a separate variable of storage type of the food in hand. This is because of the limited storage capacity in food banks and the limited options of storage types that they have.

4.3.5 Food information

The dataset consisted of the name of the food item that was provided. This data was hence used to obtain a separate variable of food types and are categorized as per the MyPlate guidelines for future reference of nutritional guidance and long-term food security improvement.

4.3.6 Time stamp information

The datasets consisted of the date in which the food was supplied to the food bank along with the day of the year, day of the week, month of the year and week of the year at which the food was supplied.

As mentioned earlier, the response variable in our analysis is the quantity of food received and purchased dataset variables are provided in Table 4.2 and Table 4.3.

Table 4.2 Summary of the key fields used in the study for Donated food dataset

Variable names	Description
Dow	Day of the week
Woy	Week of the year
Doy	Day of the year
Moy	Month of the year
Food type	Type of food (Meats, fruits and vegetables, dairy, confectionaries and bread)
Storage type	Type of storage space (Fresh produce, frozen and dry stock)
Donor ID	Unique ID of donors
Donor type	Type of donor base (Manufacturers, retailers, restaurants, others)
Quantity Received	Food supplied

Table 4.3 Summary of the key fields used in the study for Purchased food dataset

Variable names	Description
Dow	Day of the week
Woy	Week of the year
Doy	Day of the year
Moy	Month of the year
Food type	Type of food (Meats, fruits and vegetables, dairy, confectionaries, and bread)
Storage type	Type of storage space (Fresh produce, frozen and dry stock)
Source ID	Unique ID of source of purchase
Source type	Type of purchasing source (Manufacturers, retailers, restaurants, others)
Quantity Received	Food supplied

4.4 Exploratory analysis

Table 4. provides the summary of the descriptive statistics of the Quantity Received response variable for both the datasets which also highlights the heavily right-skewed tail of the response variable (as shown in Fig 4.1.). For instance, we can see that in table 4.4, for all products, 75% of Quantity has its value below 212 pounds while the maximum value is 9327 pounds. Moreover, it can be observed that for both the datasets, the maximum value is of the same range and rather high,

indicating that food banks also purchase high volumes based on necessity and funds thereby ensuring the importance of our study to include purchasing aspect as well for optimum planning operations in the organization. The distribution of the response variable for each of the datasets is depicted in Fig. 4.1 and Fig. 4.2 which tells us that a small fraction of the dataset consists of instances of disproportionately large amounts of food supply. The empirical cumulative distribution functions are depicted in Fig. 4.3 and Fig. 4.4. The y-axis shows the cumulative probability and the density plots for both datasets are steep and centered at zero, showing that large events are very rare and small events are frequent. This suggests that we should implement several modeling techniques to study the data and not just focus on a one-size-fits-all conservation policies.

Table 4.4 Summary of Quantity Received (Response Variable) for both the datasets

Parameter	Purchased Products	Donated Products
Minimum	1.0	1.0
1st Quartile	19.0	47.0
Median	50.0	101.0
Mean	392.5	218.5
3rd Quartile	173.0	212.0
Maximum	8380.0	9327.0

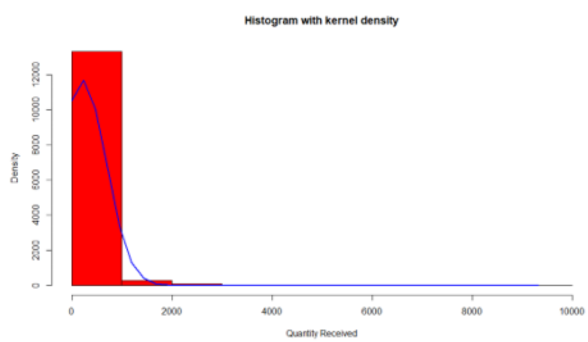


Figure 4.1 Distribution of response variable: histogram with overlain kernel density plot (Donated food dataset)

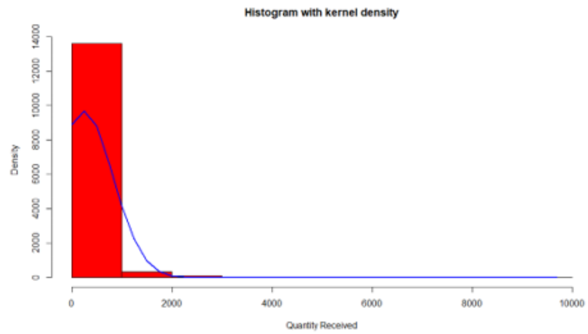


Figure 4.2 Distribution of response variable: histogram with overlain kernel density plot (Purchased food dataset)

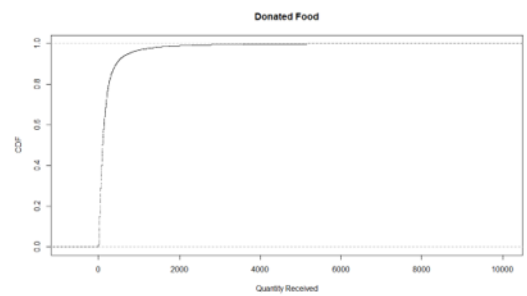


Figure 4.3 Empirical cumulative distribution functions for the study of the donated food dataset

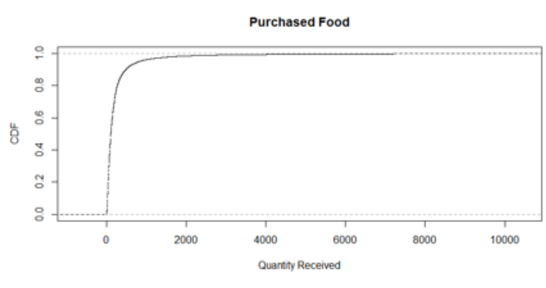


Figure 4.4 Empirical cumulative distribution functions for the study of the purchased food dataset

Since the input variables for the datasets are nominal categorical variables, observing the correlation between variables creating correlation plots does not provide the in-depth treatment of the various relationship between these variables thereby making it a possibility of reduction of optimum predictive accuracy by removing some variables based on the correlation matrix (Hastie et al., 2009). Hence, to understand how the input variables are related, the measure of association is done by developing association plots for both the datasets as shown in Fig. 4.5. And Fig. 4.6.
The measure of association does not mean causality, but rather association– whether one variable is linked to another. This association measure also indicates the relationship's strength, whether weak or strong (Agresti, 2003). The diagonal factor K in these plots denotes the number of distinct levels for each nominal variable. The forward and backward association measures for each variable pair are found in the off-diagonal components. Certain variables have a direct interaction with one another. However, the opposite relationship does not seem to be valid for the same pairs, meaning that predicting one pair from the other would be difficult.

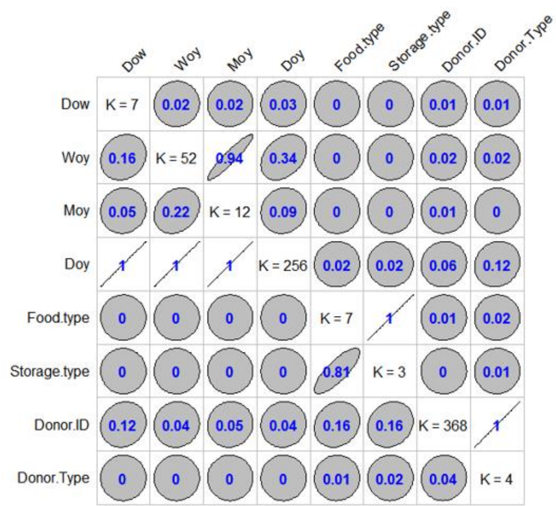


Figure 4.5 Association plot of the nominal categorical variables in the donated food dataset

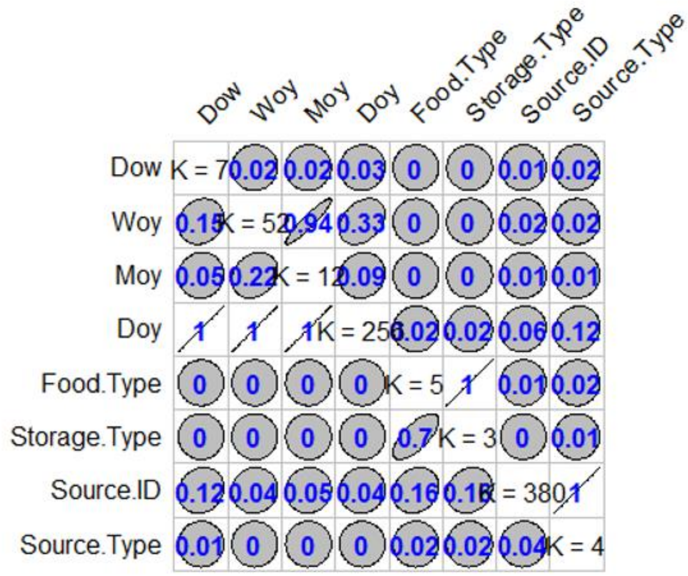


Figure 4.6 Association plot of the nominal categorical variables in the Purchased Food dataset

4.5 Methodology

This section presents the generalized research framework proposed in this study and provides a brief theoretical background of the models developed to evaluate the food supply datasets of the local food bank in Indiana.

4.5.1 Research framework

After creating the required input variables, the dataset with the collected records for each supply of food was used as the final dataset. The model development process, which is defined in the following subsections, followed this stage. Although data from the Food Finders food bank was used to illustrate the applicability of the proposed study, the approach and methodology are transferable and can be applied to other food banks and regions, as seen in this context. As shown in Fig. 4.8, the methodology developed is also available as an algorithm called the Data-driven Multi-variate Food Supply Modeling Approach.

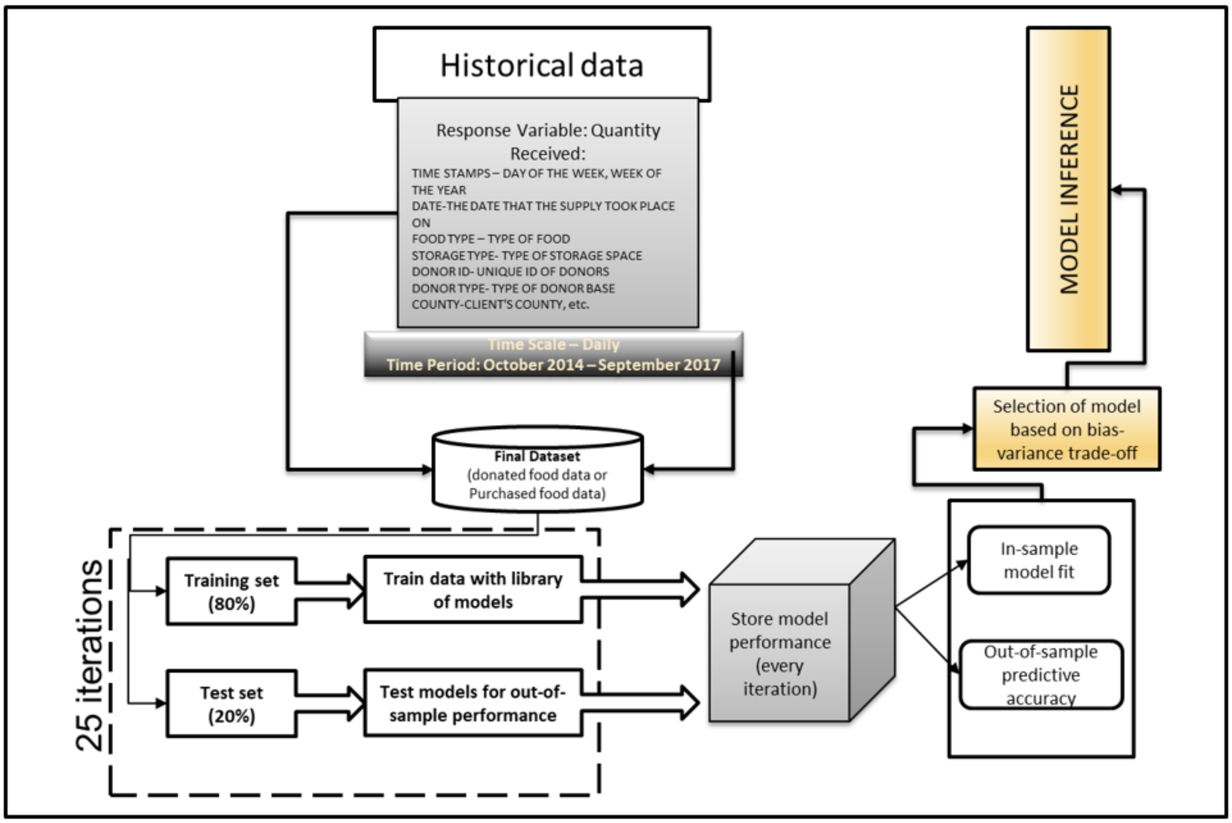


Figure 4.7 Research framework

```
Algorithm 1: Data-Driven, Multivariate
Food Supply Modeling Approach
-----------------------------------------
For a given food bank do
    identify the multivariate measures of
system
    Collect relevant data
    Leverage Machine Learning
Algorithms to estimate the distance
measures as a function of the input
parameter space
    Assess model accuracy
    if accuracy > acceptable threshold set
by the stakeholder
    else
        Improve data collection, further
model tuning
    end if
end for
```

Figure 4.8 Algorithm for predictive modeling framework developed

4.6 Methods

Several types of statistical and machine learning approaches, including parametric, semi-parametric, and non-parametric methods, have been introduced and trained on datasets containing donated foods and purchased product data. This is done in order to construct the most accurate predictive models possible, which reflect the best understanding of the dynamic and non-linear relationships between the quantity of products obtained and the various input variables (see Table

1). More exclusively, we used the methods of Generalized Linear Model (GLM), Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and Bayesian Additive Regression Trees (BART) to estimate the food supply in the given food bank organization. Based on these machine learning algorithms, predictive models of the food supply are developed employing rigorous cross-validation to highlight the model that out-performed all the others in terms of out-of-sample predictive accuracy. A succinct evaluation of each of the methods used in our study are examined below.

4.6.1 Generalized Linear Model (GLM)

GLM (Generalized Linear Models) is a linear regression extension. The normality assumption is relaxed in GLMs, enabling the response variable to be distributed according to an exponential family of distributions (e.g., Gaussian, Binomial, Poisson, Gamma, or inverse-Gaussian) and linked to the predictors through a connection feature (Cordeiro & Mccullagh, 1991; Nelder & Wedderburn, 1972). A dependent variable Y with a distribution that falls into the categories of normal, binomial, Poisson, gamma, or Inverse-Gaussian, as shown in the equations below:

$$Y_i \sim f_{Y_i}(y_i)$$

$$f_{Y_i}(y_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right\}$$

where θ and ϕ are the location and scale parameters respectively.

A set of independent variables x_i.

A link function *g(.)* binding the parameters of the dependent variable to the linear combination of input variables. GLMs are widely popular due to their ease of application and interpretability. Nonetheless, GLMs consider the ‘rigid’ assumptions of global parametric models thereby costing the performance of predictive accuracy.

4.6.2 Generalized Additive Model (GAM)

The Generalized Additive Model (GAM) is a machine learning method that is semi-parametric. It relaxed the linearity assumption used in the above-mentioned GLM process, allowing for the allocation of local nonlinearities (Hastie & Tibshirani, 1990). GAM assumes that dependent variable y has a distribution with mean $\mu = \mathrm{E}[Y|x_1, x_2, \ldots\ldots x_p]$ linked to the predictor variables via a link function as:

$$g(\mu_i) = \alpha + \sum_{j=1}^{p} f_j\,(x_j)$$

where each f_j is a smoothing function of a class of functions projected non-parametrically, like regression splines and tensor product splines.

4.6.3 Support Vector Machines (SVM)

SVM is utilized in pattern recognition by constructing hyperplanes in the given data space (Vapnik & Golowich, 1997). SVM was originally developed for classification problems but can be extended to regression of the general form $f(x) = x^T\beta + \beta_o$. The β coefficients of the regression model are estimates by the following minimization:

$$H(\beta, \beta_o) = \sum_{i=1}^{N} V(y_i - f(x_i)) + \frac{\lambda}{2}\left(||\beta||\right)^2$$

Where λ is a regularization parameter and can be estimated by cross-validation. Also, $V(r)$ is the general error measure and can be shown as:

$$v_\epsilon(r) = \begin{cases} 0, & if\ |r| < \epsilon \\ |r| - \epsilon, & otherwise \end{cases}$$

This measure makes the fitting less sensitive towards outliers and noise (Hastie et al., 2003).

4.6.4 Random Forest (RF)

RF is a tree-based ensemble data-miner that is non-parametric (Breiman, 2001). To generate the final estimate, it is a modification of bootstrap aggregation to multiple Classification and Regression Trees (CART) and taking the mean of the predictions of the roughly uncorrelated trees. The procedure consists of *B* bootstrapped regression trees (T_b) with *B* chosen based on cross-validation. Low-bias, high-variance techniques are regression trees. In other words, they can reasonably reliably assess the shape of the data (low bias), but they are extremely prone to outliers (high variance) (Hastie et al., 2009). As a variance reduction method, RF employs model averaging. As a result, the final approximation is the average of predictions from all B trees, as shown below:

$$f_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

Each tree in RF is built using bootstrapped re-samples of the input data, and split variables are chosen at random to promote tree independence. As a result, RF can make accurate predictions by lowering the correlation between the trees, enabling model averaging to generate low-bias, low-variance predictions while keeping the errors of each individual unpruned tree low.

4.6.5 Bayesian Additive Regression Trees (BART)

BART stands for Bayesian Additive Regression Trees, a non-parametric Bayesian form. To estimate the dependent variable, the BART model uses sum-of-trees. Each regression tree is constructed by recursively dividing the data area into sub-regions or nodes and tailoring a simple model (e.g., mean of the response or dependent variable) in each one. The covariates that are split and the split values that are chosen are chosen in such a way that the best fit is obtained in each sub-region. As shown in the equation (Merwe, 2018), the final model estimate is the sum of the estimates from m small trees:

$$Y = (\sum_{i=1}^{m} g(x; T_j, M_j)) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

where $g(x; T, M)$ is the function which designates the parameters of the terminal nodes of tree T, $\mu_i \in M$ to the predictors x. To ensure that each tree contributes only partially to the final predictions, regularization priors are used to control the model's complexity. As a result, regularization priors aid in reducing the influence of an individual tree's impact on the sum-of-trees model (Merwe, 2018). BART is resistant to outliers, has high predictive capacity, and is a fully probabilistic strategy, meaning it can produce complete distributions of expected response values.

4.6.6 Bias-variance tradeoff

A predictive model's ability to make good predictions on an independent test sample determines its generalization performance. The key decision-maker for minimizing generalization error is balancing the bias-variance trade-off (Hastie et al., 2009). One of the most customarily used methods for balancing bias and variance is cross-validation. To estimate predictive accuracy, we use the k-fold cross-validation process. K-fold cross-validation consists of splitting the data into *k* equal-sized subsets at random. In each duplication, the model is fitted to the subsets except the *k*th held-out sample, and the predictive accuracy is calculated based on the models' performance on the kth held-out subset. In this paper, the out-of-sample model performance was estimated using a 20% holdout cross-validation and using the following formulae:

$$MSE_{out-of-sample} = \frac{1}{k}[\sum_{k=1}^{n}\frac{1}{m}(\sum_{i=1}^{m}(y_{i,k} - \widehat{y_{\iota,k}})^2)]$$

$$MAE_{out-of-sample} = \frac{1}{k}[\sum_{k=1}^{n}\frac{1}{m}|\sum_{i=1}^{m}(y_{i,k} - \widehat{y_{\iota,k}})|]$$

k = number of times of performing cross-validation

m = hold-out numbers during every cross-validation

$y_{i,k}$ = *ith* actual observation that was randomly held-out during the *kth* cross-validation

$\widehat{y_{\iota,k}}$ = predicted *ith* observation during the *kth* cross-validation using the model developed considering the training set data during the *kth* cross-validation

Our model selection in this paper is based on both in-sample fit and out-of-sample prediction accuracies. The in-sample error is measured using the in-sample RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and adjusted R^2; while the out-of-sample error was measured using the out-of-sample RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) as discussed above.

4.6.7 Tuning parameters

Generalized Linear Model (GLM): For GLM, the value of the tuning parameters applied are listed below:

***k*:** Refers to the number of degrees of freedom used for the penalty. For our study, when k = 2 the best Akaike Information Criterion (AIC) value is obtained: k = log(n). Hence, in our study we used k=2.

Dist.= Gaussian: This parameter specifies the type of error distribution and link function to be used in the model. In this study, for all the different methods stated, we assumed that the error follows "Gaussian distribution" and the link function is taken to be an "identity" function.

Generalized Additive Models (GAM): We implemented a stepwise update methodology and cubic smoothing function which generated the best predictive accuracy among the tuning options to select the best fit model.

Support Vector Machines (SVM): The tuning parameters used for developing the SVM model are described below:

$\boldsymbol{Kernel}$: It refers to the type of hyperplane used to separate the data. In our research, we use the "radial" kernel considering our data is non-linear.

$\boldsymbol{Gamma}$: This hyper parameter accounts for the smoothness if the decision boundary and manages the variance of the model. If Gamma is large, then we get fluctuating decision boundaries leading to overfitting. If the Gamma is small, the boundary is smoother and has low variance. The default value of Gamma is used.

Random Forest (RF): The tuning parameters for the RF model are considered below.

$\boldsymbol{mtry}$**:** The number of variables randomly sampled as candidates at each split while growing the trees.To be noted that the default values for the regression tree is $p/3$, where p is the number of independent variables used in the model.

$\boldsymbol{ntree}$: This parameter refers to the number of trees to grow. This must not be set to a very small number to guarantee that every input row will get predicted at least a few rounds. In our research, we selected the value of the $ntree$ that yielded the least mean square error (mse) while growing the trees.

Bayesian Additive Regression Trees (BART): The tuning parameters used in the BART model are described below:

***k*:** For regression, k determines the prior probability that $E(Y|X)$ is contained in the interval $(y\_\{min\}, y\_\{max\})$, based on a normal distribution. For example, when we have $k = 2$, we get the prior probability to be 95%. For classification, k determines the prior $E(Y|X)$ between $(-3,3)$. Note that a larger value of k results in more shrinkage and a more conservative fit.

***nu*:** It refers to the degrees of freedom for the inverse χ^2 prior.

***q*:** This parameter refers to the quantile of the prior on the error variance at which the data-based estimate is assigned. It is to be stated that greater the value of q, the more forceful is the fit; this is because it corresponds to placing more prior weight on values lower than the data-based estimation . It is not used for classification.

***m*:** This parameter refers to the number of trees to be grown in the sum-of-trees model.

4.7 Results

To investigate the sensitivity of food supply to different parameters, we built predictive models for data consisting solely of donated foods and data consisting solely of purchased food items. We used Generalized Linear Models (GLM), Generalized Additive Models (GAM), Support Vector Machines (SVM), Random Forest (RF), and Bayesian Additive Regression Trees to train each dataset (BART). In this section, we will go through the output of each of the trained models before looking at the significant predictors as model inference based on the final model that outperforms them all in terms of out-of-sample predictive accuracy and goodness-of-fit.

The model results for both the datasets (Table 4.5 and Table 4.6) indicated that RF provides substantial results as compared to all other tested statistical learning methods and GLM demonstrated the poorest performance in comparison to all the other models in both the cases. This can also be substantiated by the goodness of fit (R^2) value and predictive accuracy (out-of-sample RMSE and MAE) of RF method for both the datasets. Based on the goodness of fit, and predictive accuracy (out-of-sample RMSE and MAE). We also observe that using our proposed predictive method, RF can assist food banks in more inferencing of influential factors as well as better replicability of our study due to its easy inferencing function and ease of tuning parameters since it is a non-parametric model. This further supports our hypothesis that linear models do not adequately represent the non-linearities in the results. Furthermore, in both cases, RF and BART are viable options. However, we chose RF as the best model for predictive study for our datasets because of the bias-variance trade-off and the potential to study model inferences effectively. The percentage improvement yielded by each of the trained models over having no statistical model and using the historical average as a predictor (i.e., the 'mean-only' model) is provided in Table 4.7 and Table 4.8 for donated and purchased food datasets respectively.

Fig. 4.9 and Fig. 4.10 shows the plot of predicted versus observed values of quantity received for both donated food and purchased food products. It can be seen from the figure that while our models estimate the lower ends of quantity received well, they tend to underestimate the more extreme ends of food products for both the cases. Since the method of random forest is a non-parametric machine learning method, inference is implemented by ranking the input variables based on their contribution to out-of-sample predictive accuracy of the predictive model. Fig. 4.11 and Fig. 4.12 shows the ranking of the predictors. In Fig. 4.11 and Fig. 4.12, the highest variable on the y-axis corresponds to most important variable, and the x-axis shows the level of predictive accuracy reduced if the variable is removed. As seen in Fig. 4.11 and Fig. 4.12, Donor/Source ID, Day of the year, donor type, food type and week of the year variables are identified as the topmost important predictors in both the datasets. Thus, it can be concluded that the donor base and source base, the food information and the time stamps are the most important predictors of food supply data whether it is donated or purchased. Fig.4.11 shows that food type becomes a more important predictor for donated foods than for purchased food dataset. Moreover, it can also be seen quite intuitively that based on the week supply of food, the food is purchased as the planning of food takes place on a weekly basis in the food banks, the importance of week of the year variable above food type for all food data becomes apparent. This leads to future scope of studying the relation between donated foods and purchased foods and observing the correlation and causation of each of the datasets to further improve with predictivity and forecasted accuracy.

The results obtained from our predictive modeling study has helped to characterize the various factors that affect the supply of food behavior in the food bank system. We believe that the supply of both donated and purchased food products estimated in our study can be considered as initial information by the decision makers and planners in the organization to design optimal routing models with objectives of equitability and effectiveness in their food distribution efforts owing to our high predictive accuracy results and repeatable steps based on our predictive framework.

Table 4.5 Donated food model performance results

Sl. No.	**Model**	**Tuning Parameters**	**R^2**	**In-sample**		**Out-of-sample**	
				RMSE	**MAE**	**RMSE**	**MAE**
1	Mean (Null model)	-	-	-	-	523.55	489.29
2	Generalized Linear Model (GLM)	k=2.0, Dist.=Gaussian	0.8977	224.11	59.57	231.23	71.23
3	Generalized Additive Model (GAM)	Stepwise update	0.8793	230.0937	9.6586	238.25	11.256
4	Support Vector Machines (SVM)	Kernel = Radial; Gamma = scale	0.8845	221.11	58.88	230.59	68.52
5	**Random Forest (RF)**	**mtry=p/3 =3; ntree=200**	**0.9136**	**189.72**	**49.45**	**194.56**	**52.35**
6	Bayesian Additive Regression Trees (BART)	k=2,nu=10,q=0.75,m=200	0.9089	194.34	65.34	201.12	69.35

Table 4.6 Purchased food model performance results

Sl. No.	Model	Tuning Parameters	R^2	In-sample		Out-of-sample	
				RMSE	MAE	RMSE	MAE
1	Mean (Null model)	-	-	-	-	624.33	598.93
2	Generalized Linear Model (GLM)	k=2.0, Dist.=Gaussian	0.7823	222.96	66.38	240.87	71.65
3	Generalized Additive Model (GAM)	Stepwise update	0.7768	231.01	9.86	257.34	12.73
4	Support Vector Machines (SVM)	Kernel = Radial; Gamma = scale	0.8033	210.87	67.34	231.03	72.14
5	**Random Forest (RF)**	**mtry=p/3 =3; ntree=180**	**0.9762**	**172.78**	**52.65**	**187.56**	**61.76**
6	Bayesian Additive Regression Trees (BART)	k=2,nu=3,q=0.75,m=50	0.8546	169.78	67.99	196.92	72.22

Table 4.7 Percentage improvement over the 'null' model for donated food model performance results

Models	Out-of-sample error	
	RMSE	MAE
GLM	0.558342088	0.854421713
GAM	0.544933626	0.976995238
SVM	0.559564511	0.859960351
RF	**0.628383153**	**0.893008236**
BART	0.615853309	0.858264015

Table 4.8 Percentage improvement over the 'null' model for purchased food model performance results

Models	Out-of-sample error	
	RMSE	MAE
GLM	0.614749772	0.877458857
GAM	0.617196872	0.979314311
SVM	0.629955312	0.879551868
RF	**0.7019017**	**0.893966815**
BART	0.680773475	0.882752905

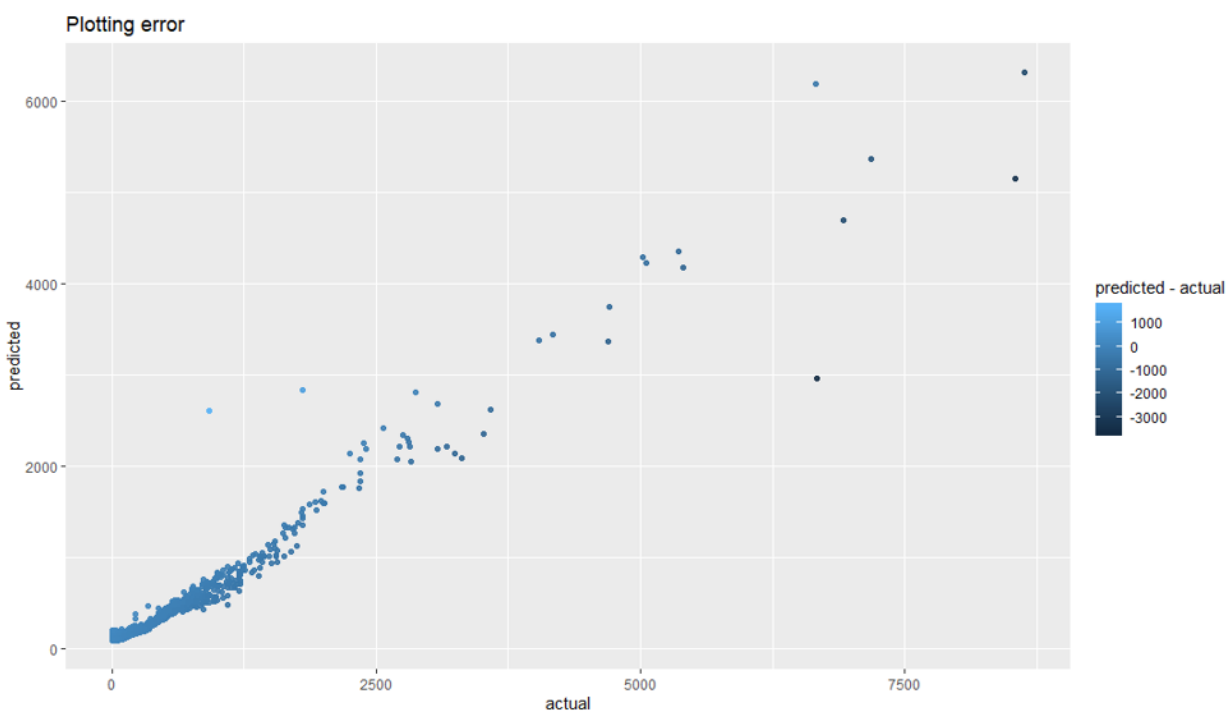


Figure 4.9 Plot of predicted versus observed for donated food data

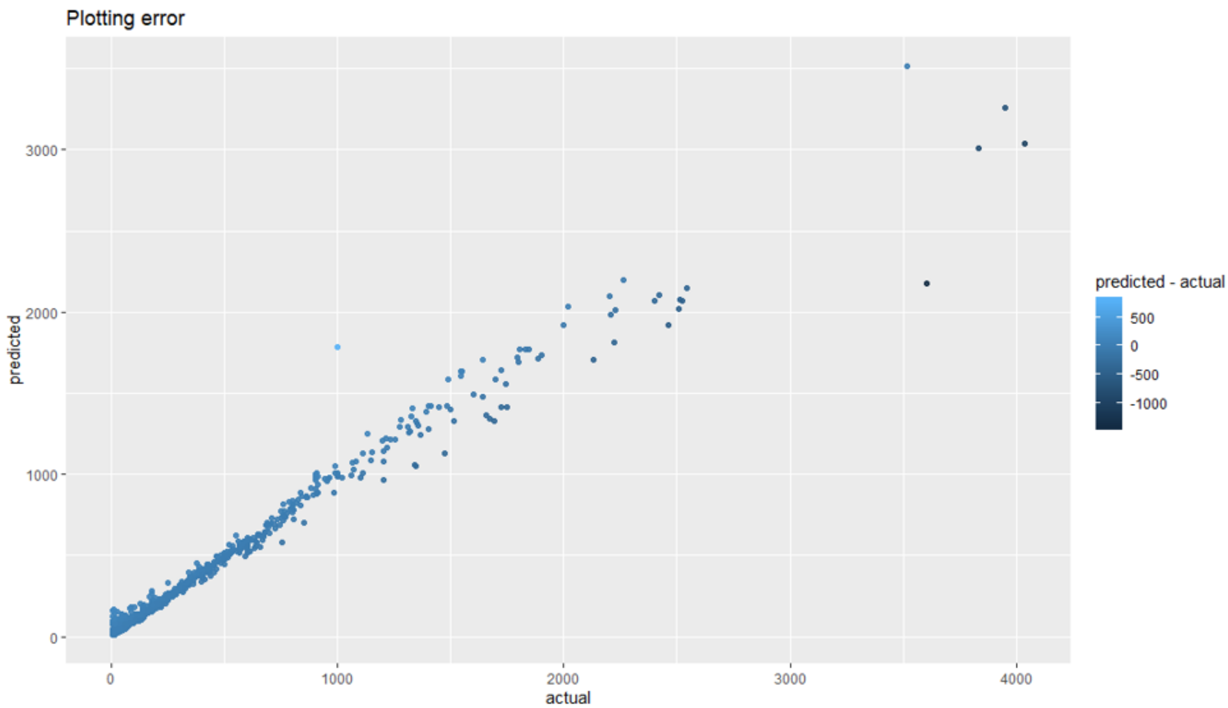


Figure 4.10 Plot of predicted versus observed for purchased food data

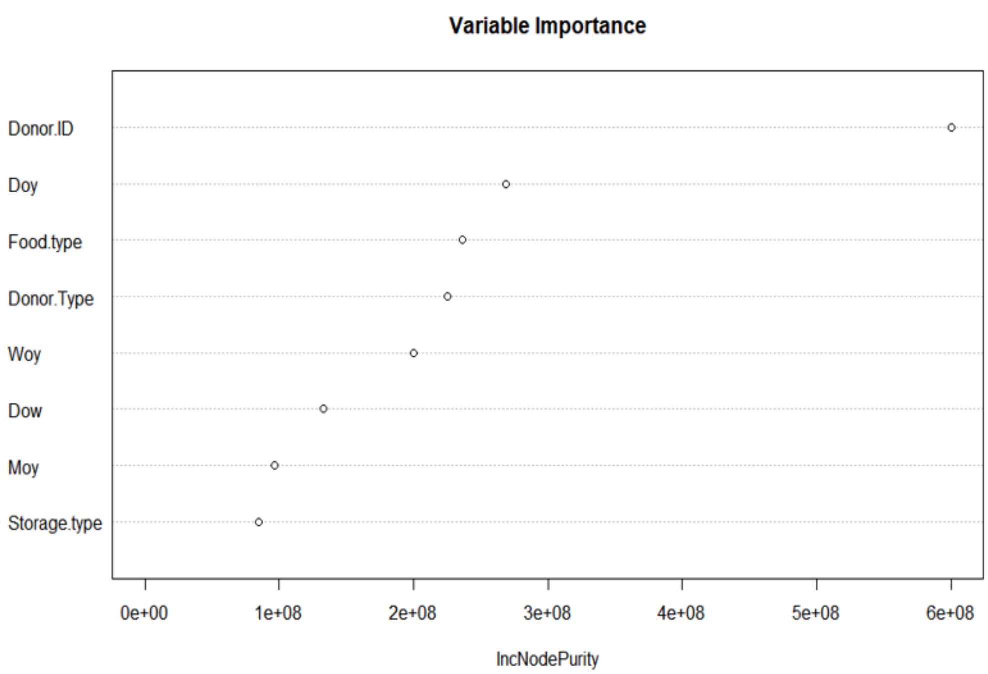


Figure 4.11 Variable importance plot: Donated food data

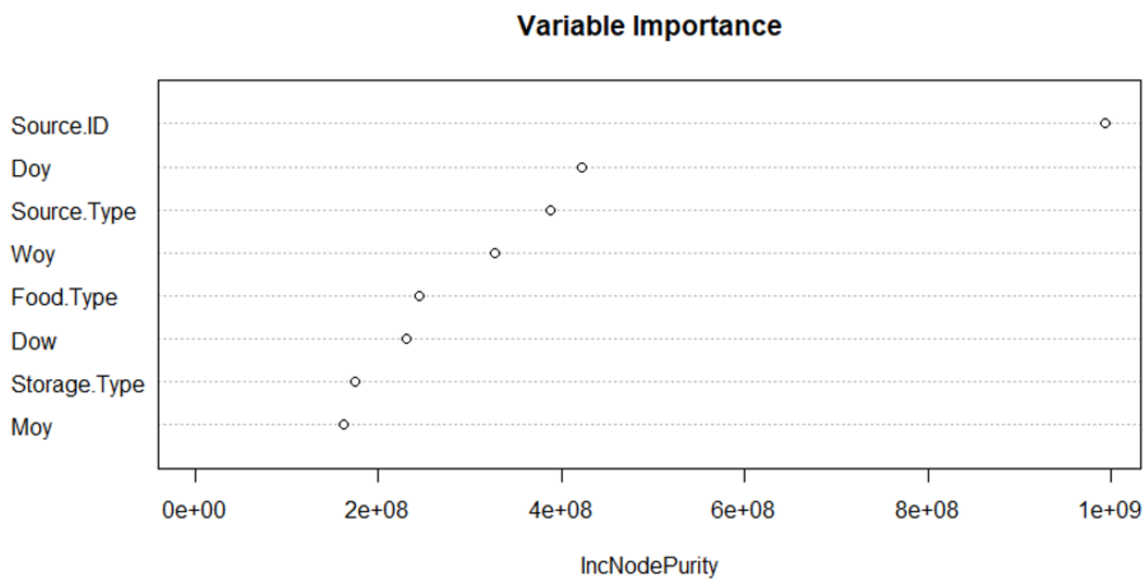


Figure 4.12 Variable importance plot: Purchased food data

4.8 Discussion

4.8.1 Proposed predictive framework implementation on use case studies: an observation

Our research framework of implementing predictive modeling in the food supply aspect of food bank logistics can be used across different planning projects and situations such as budgeting, facility location, routing, cost reduction and optimization. Predictive analysis aids in every stage of planning namely- strategic, tactical and operational. In order to validate the applicability of the proposed framework for each stages of planning, its implementation is provided in three case studies: Vehicle routing of a food bank system (operational planning case), Warehouse costing and improvement (tactical planning case), and Budget planning of a food bank organization (strategic planning case). Case 1 analyzes the procedure for budget planning, while case 2 analyzes the space planning and costing of a warehouse owned by a food bank and finally, case 3 analyses the vehicle routing problem for a single food bank and its respective food agencies. We provide the implementation of the results of the proposed framework onto these planning problems and present predicted results in each of the case studies to demonstrate the potential of the proposed framework and its easy implementation in achieving optimal planning performance at any level (Operational, Tactical or, Strategic)

4.8.2 Case 1: Strategic Level Planning- Predictive budgeting for food banking

Effective management of finances is critical to organizational success. Budgeting will outline the major costs and give an overview of available capitals. Monetary Donors also find it useful when tracking their contributions to see how their funds are being used. Having a wide-ranging budget will establish integrity with the donors and provide a clear view of goals that can be set for the following year (Gutjahr & Fischer, 2018). The strategic planning report of a food bank in literature is used as a case study (Second Harvest Food Bank of Central Florida, 2020). In the report, the budgets are drawn to cover a fiscal year and must be made ready before the beginning of each year. To set reasonable projections of financial need, it is important to have accurate forecasts and analysis. Based on the needs of budgeting, the proposed predictive modeling framework and methodology (Section 4.4) using Random Forest is implemented to provide the current food resources and the predicted food resources for displaying the accuracy of the forecasts and finally provides a forecast of the coming year food resources. Since budgets are developed annually, Table 4.9 and Table 4.10 provides the predictive results on an annual basis.

Table 4.9 Poundage of donated foods on an annual basis

Food donor type	Poundage in 2017 (Actual)	Poundage in 2017 (Forecasted)	Variance (%)	Poundage in 2018 (Forecasted)
Manufacturers	345615	360130	4.2	352478
Retailers	1766244	1823646	3.4	1976618
Restaurants	150803	155779	3.2	160030
Others	255133	265889	4.9	230882

Table 4.10 Poundage of purchased food on an annual basis

Source type	Poundage in 2017 (Actual)	Poundage in 2017 (Forecasted)	Variance (%)	Poundage in 2018 (Forecasted)
Manufacturers	354841	367615	3.6	345239
Retailers	411261	428122	4.1	397374
Restaurants	51765	50315	2.8	52347
Others	100573	104495	3.9	109384

From the results mentioned in Table 4.9 and Table 4.10, we can see that the variance of the predicted values to the actual values is less than 5% which is beneficial in the budget planning as it is recommended to ensure variance is less than 10% for efficient budgeting (Shim et al., 2011).

4.8.3 Case 2: Tactical Level Planning – Warehouse costing and improvement of a food bank system

As mentioned in Section 1, food banks distribute and provide food to their respective food agencies with the food insecure people receiving food by visiting these food agencies. To maintain the food products donated and handle the growing food distribution and growing food demand, the food banks need to ensure optimal warehouse spacing and inventory management (Shrestha, 2009). There has been previous research conducted to improve the warehouse spacing and inventory management which specifies the need for optimal food handling with the help of effective predictive modeling. The usage of the proposed predictive modeling framework ensures the success of accomplishing this challenge faced by the tactical planners of the food bank system. Using the proposed framework, we can determine the peak months of inventory and visually provide aid to the management team and understand their load of inventories and prepare for handling them in the future. This can also aid in the decision making of understanding the need for warehouse remodeling and investment of reconstruction if need be. Table 4.11 and Table 4.13 below provides the actual quantities of food based on the storage type variable in the dataset and results for prediction are extracted on a monthly basis for the upcoming period for both donated and purchased food datasets and provided in Tables 4.12 and Table 4.14.

Table 4.11 Monthly food poundage of donated foods (2017)

Storage Type	January	February	March	April	May	June	July	August	September	October	November	December
Dry stock	117852	96448	97217	117206	109501	96645	127597	109970	98692	103404	162763	126770
Fresh Produce	58496	58962	80318	85920	136448	113701	132010	198496	159361	53500	41920	72211
Frozen	104826	79413	90793	67028	63417	54405	57169	51811	51020	94646	86944	120662

Table 4.12 Monthly food poundage of donated foods (predicted for 2018)

Storage Type	January	February	March	April	May	June	July	August	September	October	November	December
Dry stock	165135	244903	223808	257883	293885	240165	169532	201287	200647	176367	231203	194121
Fresh Produce	30932	31922	43291	62149	99605	60437	110781	105224	92935	125900	109917	44559
Frozen	215591	103679	105647	87704	100351	97867	53031	135269	146040	81138	55813	99230

Table 4.13 Monthly food poundage of purchased foods (2017)

Storage Type	January	February	March	April	May	June	July	August	September	October	November	December
Dry stock	222488	167154	187509	115663	105237	133549	131400	207665	162179	207156	209118	186279
Fresh Produce	22904	41471	44453	37062	33381	43290	84967	91201	78389	98990	45499	79579
Frozen	78018	70098	192392	72632	70215	64728	52087	90092	115912	148448	81854	151306

Table 4.14 Monthly food poundage of purchased foods (predicted for 2018)

Storage Type	January	February	March	April	May	June	July	August	September	October	November	December
Dry stock	204217	254206	186223	190566	298188	139977	194424	220641	204179	164579	150151	175037
Fresh Produce	116091	49222	148898	76587	91026	64628	112093	124673	117414	86285	41607	98117
Frozen	186497	109986	176377	116320	180681	101771	94273	85968	93367	113846	46580	73263

4.8.4 Case 3: Operational Level Planning- Vehicle routing of a food bank system using predictive modeling results

For food rescue and delivery, vehicle routing is a very complex problem that includes challenging aspects like, uncertainty in donations and donors, cost-effective scheduling of pickup and delivery nodes, cost-effective routing, limited transport resources, perishability of the rescued food equitable distribution of rescued food, etc. (Nair et al., 2016a). Implementing forecasting models to handle uncertainty in supply of donated and purchased foods helps in the better planning of scheduling and routing models. Vehicle routing has been a central component in the logistics of surplus food rescue and delivery operations, where the decision makers have to consider multiple criteria such as transportation cost, customer service requirements, operational constraints, etc.

To effectively schedule the visits of food providers and welfare agencies, we provide case instances which are based on our proposed framework that utilizes real-life information and results that are of good accuracy percentages that ensure the right direction towards effective and efficient operational planning. Since we have considered Donor ID as one of the independent variables in our study, we are able to extract predictive results right from the Donor ID level which can be used for operational planning purposes. Table 4.15 provides a case instance sample for a single day that has been extracted using the framework proposed and can be easily implemented to develop case instances on a daily basis for vehicle routing optimization of food bank network.

Table 4.15 Case instance for a generic vehicle routing planning and scheduling for a single day time horizon

ID	Lat	Long	Bread and Cereal	Fruits and Vegetables	Meats and Proteins	Dairy Products	Confectionaries and others
1	41.452	-81.691	116	60	122	350	138
2	41.458	-81.722	114	57	172	152	177
3	41.497	-81.668	110	52	200	294	178
4	41.506	-81.651	178	52	178	178	184
5	41.452	-81.549	120	57	193	328	111
6	41.398	-81.51	155	60	91	333	165
7	41.607	-81.527	119	51	102	391	185
8	41.574	-81.578	193	54	176	171	170
9	41.457	-81.769	115	59	198	237	126
10	41.492	-81.654	172	55	129	347	160
11	41.525	-81.438	173	55	174	397	197
12	41.487	-81.778	186	54	132	398	171
13	41.52	-81.65	134	56	96	174	194
14	41.513	-81.62	180	58	160	149	169
15	41.538	-81.587	139	51	158	282	164
16	41.533	-81.62	189	60	102	351	103
17	41.525	-81.617	101	53	181	389	152
18	41.537	-81.613	109	54	138	129	121

4.9 Conclusion and Future Directions

In this paper, the supply of different types of food both donated and purchased is investigated using advanced machine learning algorithms. A generalized data-driven framework is proposed to assess the supply estimate and to identify the key predictors in the model. Although the proposed model is used to characterize the food supply data, it can also be leveraged for varying time durations (week-based, month-based, etc.) as it outperforms all the other models by means of explaining the variations in data as well as out-of-sample predictions.

This paper evaluates the impact of five parametric, semi-parametric and non-parametric machine learning techniques: GLM. GAM, SVM, RF, and BART to explore the patterns in the supply of donated food and purchased food process in the food bank organization. The results suggest that Random Forest (RF) outperforms all other tested statistical methods in terms of capturing the supply estimate for both donated and purchased goods. Food donation and purchase of food takes place based on a function of many factors. The study identified the development of new variables namely, the type of foods being available at the food bank. The types were classified based on the MyPlate guidelines provided by the USDA and the development of variables based on storage and donor base. We also showed estimate results for data consisting of purchased food products indicating the importance of this aspect in the supply study of food banks which has been neglected in the literature. These variables have a significant non-linear impact on food supply, and we were also able to identify the key predictors in the study and were able to infer from these predictors the donation behavior and the purchasing habit of food banks. While the model developed provided satisfactory results, it would be beneficial if further characteristics of the region are included in the study (e.g., demographic details, economic metrics, etc.) which is one of the ways in which this work can be further extended. Steps can be taken to study the causation and correlation between the two separate datasets (donated foods and purchased foods respectively) to further strengthen the forecast accuracy.

Another area of future research is the involvement of these predictive models with optimization models such as routing and inventory management to ensure optimal allocation of food to the people in need. Additionally, also considering predicting the demand of food that the food banks receive and providing similar predictive study for the same would be suitable to ensure efficient and fair distribution of food in the food bank network.

In closing, we would like to point out that our analysis was focused on ensuring a wholesome study of the supply of food in the food bank network that has been segregated and neglected in the food bank literature. We believe that having a predictive study is a key first step in devising effective policies of increasing food security and sustainability with the added advantage of minimizing food waste and evaluating the environmental impact that food banks bring to the nation.

5. GMM CLUSTERING FOR IN-DEPTH FOOD ACCESSIBILITY PATTERN EXPLORATION AND PREDICTION MODEL OF FOOD DEMAND BEHAVIOR

5.1 Introduction

Food insecurity is branded by the uncertainty or absence of ability to acquire nutritionally satisfactory and safe nourishment in ways that are socially acceptable (e.g., without resorting to stealing, rummaging, or different sorts of adapting strategies). This condition is unavoidable, influencing masses everywhere throughout the world. In the United States, it impacts every community with food insecurity existing in each region in America (Feeding America, 2015). The United States Department of Agriculture (USDA) has evaluated that starting at 2017, nearly 40 million individuals have been living in sustenance-uncertain household units with 6.5 million of them forming minors (Feeding America, 2015). In the United States, there are a range of assistances connecting collective efforts between government, public and private bodies for food insecure individuals. One of the largest national non-profit hunger relief organization tackling hunger and food insecurity in the country is Feeding America.

There are around 200 food banks and close to 60,000 food agencies (Consisting of food pantries, meal programs, etc.) that are being spearheaded by Feeding America to offer food and assistance to the food insecure people and households. Food and donations are mostly recovered from national food and grocery producers, suppliers, shippers, packers, and farmers, as well as public bodies and other organizations, and transported to food banks, which serve as food storage and delivery depots for smaller front-line food agencies (Feeding America, n.d.). The under-privileged population can, hence, receive donated foods from these food agencies. In general, in this donated food supply chain network, there is high uncertainty in the supply and demand of donated foods for the food insecure. The demand in specific, is usually hidden to the food bank as previously, the charitable food agencies have been providing little to sometimes, inconsistent information to the food banks thereby causing inefficient food distribution to these agencies further affecting the levels of food insecurity in these regions.

The final goal of a bequest-driven supply chain such as the food bank supply chain is to maximize the relief for the people in need while minimizing food waste as a by-product benefit. We fill this gap by explicitly studying the nature of the families visiting the food agencies based on demographical information and define food assistance deserts in the given region of study by applying concepts of unsupervised machine learning (Gaussian Mixture Models (GMM) clustering) to observe the geographic and demographic intricacies of the given region in detail. Once the dataset is examined and analyzed comprehensively, the results obtained from clustering the dataset is used for forecasting demand-side inputs using supervised machine learning forecasting techniques. A number of studies in the public policy and health literature examine the usage of food banks and the challenges associated with limited and unpredictable supply, and the forecasting of supply uncertainty (Nair et al., 2017a). However, to the best of our knowledge, the application of statistical analysis techniques to handle the demand uncertainty in the food bank supply chain system has not been addressed. We study, the nature of the current food insecure household situation in the given region using unsupervised machine learning methods such as clustering and demonstrate that we can get reasonable estimates for demand of food by implementing supervised machine learning techniques on the clustered data. Our results generate forecast accuracy of 82% for specific instances.

Our study has particular merit because it is important for non-profit organizations to leverage knowledge and technology to renovate and reinvent their preparedness and effectiveness. Food banks having information of their current and future demand behavior can help improve their food distribution efficiency and hence, make suitable informed distribution decisions, thereby meeting with their objectives of ensuring effective, equitable, and efficient food-aid operations and distributions to the people in need.

The remainder of the paper is organized as follows- Section 5.2 provides a summary of the literature. Section 5.3 presents the data and methods, and Section 5.4 summarizes our results, followed by our conclusions in Section 5.5.

5.2 Literature Review

Food insecurity and hunger are termed as long-term humanitarian issues, requiring the need to consider the necessity for evenhanded dispersion of resources (Orgut et al., 2018). There has been broad research done in the area of humanitarian logistics with significance towards the issues and challenges faced by non-profit food assistance programs such as food banks and food pantries as enlightened by Davis et al. (2014). Food bank supply chains line up with the description of humanitarian supply chains by responding to the disaster of food insecurity which can occur unexpectedly (i.e., job loss, natural calamity, etc.) or slowly (i.e., destitution) (Beamon & Balcik, 2008).The research presented in this paper aids in finding the different possible factors affecting food insecurity and hunger, thereby facilitating in improving the accessibility of food and equitable distribution of resources to the people in need. According to Waity (2016), there are innumerable ways of studying the food accessibility for people. One among these methods is considering food deserts. Food deserts are regions deficient in sources of healthy nutriments. This concept of food accessibility will be implemented in our paper.

There has been a decent amount of work that studies the subject of food bank supply chain. Mathematical models were presented by (Orgut et al., 2016a; Orgut et al., 2017) to enable the equitable and effective distribution of food donations to the people in need. Linear programming models were formulated with the maximization of effectiveness and an equity constraint developed to solve the distribution of donated foods. Deterministic network-flow models were used to reduce the quantity of undistributed food. Several logistical issues that are being faced by the food banks have also been taken into consideration by considering the transportation schedules and permitting food banks to gather food from the limited food donors and finally transporting it to the food agencies (Davis et al., 2014). In this paper, Food Delivery Points (FDPs) were proposed. FDPs were obtained by locating them using geographical information. The vehicle capacity and food degeneration constraints were considered during the assignment of food agencies to the respective FDPs. Using the optimal assignment, schedules were created that reflects the collection and distribution of donated food. However, these mathematical models do not investigate the varying demands of the various food agencies from where the accessibility of food is studied.

Demand of food from the food pantries and other food agencies has been taken as a deterministic value in previous non-profit based supply chain literature. On the hindsight, the demand of food is dynamic and uncertain in nature. Obtaining a way to observe the demand of food from the food agencies that the food banks are assigned to aids in understanding the possible issues arising out of food insecurity. According to Beamon and Balcik (2008), demand comes in the form of supplies and people for non-profit organizations and in the form of products and services for for-profit organizations, with varying demand patterns for both. Sucharitha and Lee (2018) developed a food distribution policy using suitable welfare and poverty indices and functions to ensure an equitable and fair distribution of donated foods as per the varying demands and requirements of the people. However, the factors causing the demand or food insecurity had not been considered thereby taking several assumptions in their simulation study. Also, the model developed was suitable only for a single day period. Supply of the donated food can be done based on suitable forecasting procedures. Supply based on this kind of non-profit supply chain would be mainly dealing with guaranteeing enough inventory for the demand and reviewing the changing nature of the supply of the different types of donated foods.

In terms of implementing suitable data mining techniques, there has been relevant literature discussing the role of these techniques in the estimation of future supply using historical data in various domains. In terms of using forecasting techniques in estimating the dynamics of food donation and distribution process, Davis et al. (2016) performed comprehensive numerical studies to quantify the extent of uncertainty in terms of the food donors, the food products, and the supply chain structure. Several predictive models were developed to estimate the quantity of in-kind donations. Predictive modelling techniques like multiple linear regression, structural equation modelling and neural networks were used in Nair et al. (2017) to study the dynamics of food donation behavior thereby, predicting the daily average food donated by different food providers in the given region. However, the lack of statistical analysis techniques used for the study of the demand dynamics in the food bank supply chain has been evident and has been mentioned as an important challenge from the non-profit supply chain perspective (Orgut et al., 2016b). Recent work addressing this issue has aided in the better understanding for the mitigation of food insecurity. Alotaik et al. (2017) implemented K-means clustering to identify the food assistance deserts, a term coined by Waity (2016) while analyzing the spatial inequality existing between the rural and urban areas in access to food agencies. The results obtained from the analysis by Alotaik et al. (2017) was useful in targeting the underserved areas in the given region. Finding trends and detailed observations is possible using unsupervised learning methods such as clustering which is not the case when spatial analysis is implemented (Waity, 2016). However, considering the dataset used consists of variables of different sizes and density, the affected families and certain traits could have been hidden and unobserved keeping in mind the lack of flexibility in a clustering technique such as K-means clustering (Verma et al., 2012). Implementing a soft clustering method such as GMM in our study, ensures a better visibility of traits and hidden features in a dataset featuring spatial and demographic information as compared to K-means clustering (Wang et al., 2019).

In this paper, we address the issue of food insecurity in Ohio by analyzing the food agency service data provided by Greater Cleveland Food bank (GCFB), Ohio and combining the demographic data provided by the USDA and implementing GMM clustering method to the combined data based on the distances between the family visiting the food agency and the food agency serving them and observe the factors leading to food insecurity and provide ways to increase the accessibility of food. The clustering results obtained is then implemented for the food demand predictions by developing predictive models using various statistical learning methods for the dataset modified to contain the quantity of people visiting the food agency as the response variable. We assessed the model's performances, both with clustering results and without clustering results based on their predictive accuracy to select the best model based on both generalizability and ability to capture structure of the data.

5.3 Data

5.3.1 Data Description

Greater Cleveland Food Bank (GCFB) provided service data of all the food agencies that obtain food from the distribution methods carried out by them. The study area is shown in Fig. 5.1 with the food agencies depicted in red and the families visiting these agencies depicted in grey. The plot was developed based on the dataset values of the latitude and longitude variables of the food agencies and the families visiting these organizations. GCFB distributes to food agencies situated in several counties in Ohio. The service data consists of 600,000+ records for the fiscal year of 2018 where each row represents one service to one family.

It includes the latitudinal and longitudinal location points of each family and the represented food agency that they have visited during that period of time. It was imperative to obtain the distance between each family and their visited food agency to observe if they are located at an acceptable distance or not. The dataset entailed the census tract and census block details of each family. Hence, along with the census details, the distance between the family and the food agency is also aligned and tallied. Since the region under study is predominantly an urban area, the threshold level of distance to be considered as a food assistance desert or low food accessibility is taken to be a 1-mile demarcation (Mattogno et al., 2014). The GCFB service dataset also provided specifics of the number of children, seniors and adults in every family that is being served by their food agencies. USDA also provides information of the household income and median household income of the counties at the census tract level. This data was obtained to study the low- and high-income population for the region of observation and aggregated to the service dataset as well (see Table 5.1).

For the predictive modelling study, there has not been any available study for the estimation of food demand in the food bank network. To develop this study, the aggregated dataset compiled for the clustering study undergoes data wrangling (Xu & Tian, 2015) to obtain the response variable of the total number of people visiting a particular food agency along with the time stamps, and geographical information of the person as input variables for the predictive modelling. This wrangled dataset consists of 15000+ records for the fiscal year of 2017-2018. By doing this we can predict the number of people visiting a food agency and obtain the amount of food that a food agency would require to satisfy the demands of the clients on a daily, weekly or monthly basis. For this study, we consider daily demand. However, the dataset can be modified with ease to study the weekly and monthly food demand.

We will term the former dataset developed for clustering analysis as the aggregated dataset and the dataset obtained from wrangling the aggregated dataset as the data of food demand from now on.

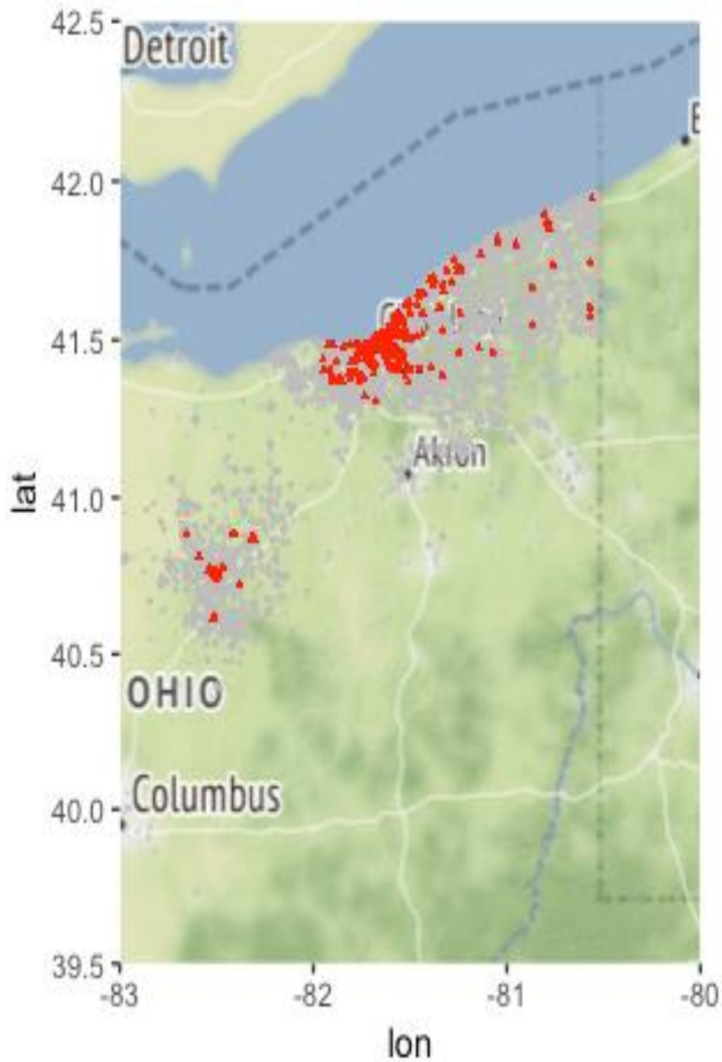


Figure 5.1 Region of Ohio (Study Area)

Table 5.1 Aggregated dataset

Attribute	Description
Date	The date that the service took place on
Family ID	System generated ID number for each client
City	Client’s city
State	Client’s state
Zip	Client’s zip
County	Client’s county
Count Adult	Number of family members between ages of 18-59
Count Child	Number of family members below age 18
Count Senior	Number of family members age 60 or above
Agency Number	ID of the food agency
Family Latitude	Client’s latitude, obfuscated to three decimal places
Family Longitude	Client’s longitude, obfuscated to three decimal places
Agency Latitude	latitude of agency/pantry
Agency Longitude	longitude of agency/pantry
County Income	Household income (median) in a given county

Table 5.2 Food demand dataset

Variable names	Description
Dow	Day of the week
Woy	Week of the year
Doy	Day of the year
Moy	Month of the year
Agency Number	ID of the food agency
Food Demand (No. of people)	Total number of people visiting the food agency

5.3.2 Exploratory analysis

The summary of the descriptive statistics of the food demand dataset by means of the number of people visiting the food agency on a daily basis is provided in Table 5.2. From the table, we observe that 75% of the overall demand has its value below 108 count while the maximum count is 5588. From observing the distribution of the response variable in Fig 5.2, we see the heavy tail of the response variable. The empirical cumulative distribution function plot in Fig 3 suggests that a small fraction of the response variable includes a huge number of people visiting the respective food agency. The y-axis shows the cumulative probability and the density plots for both datasets are steep and centered at zero, showing that large events are very rare and small events are frequent This implies that we need to estimate the demand by considering several modelling techniques to study the data and not just focus on a one-size-fits-all approach.

Table 5.3 Summary of Food Demand (Response Variable)

Parameter	**Food Demand (Number of People)**
Minimum	1.0
1st Quartile	19.0
Median	51.0
Mean	114.9
3rd Quartile	108.0
Maximum	5588.0

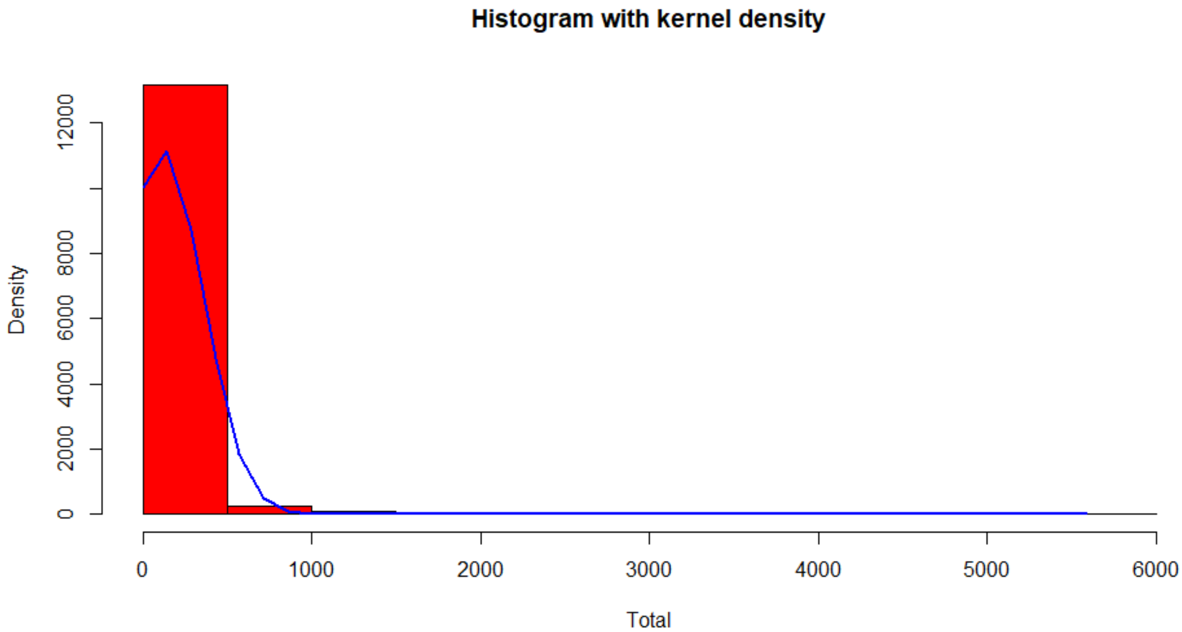


Figure 5.2 Distribution of response variable: histogram with overlain kernel density plot

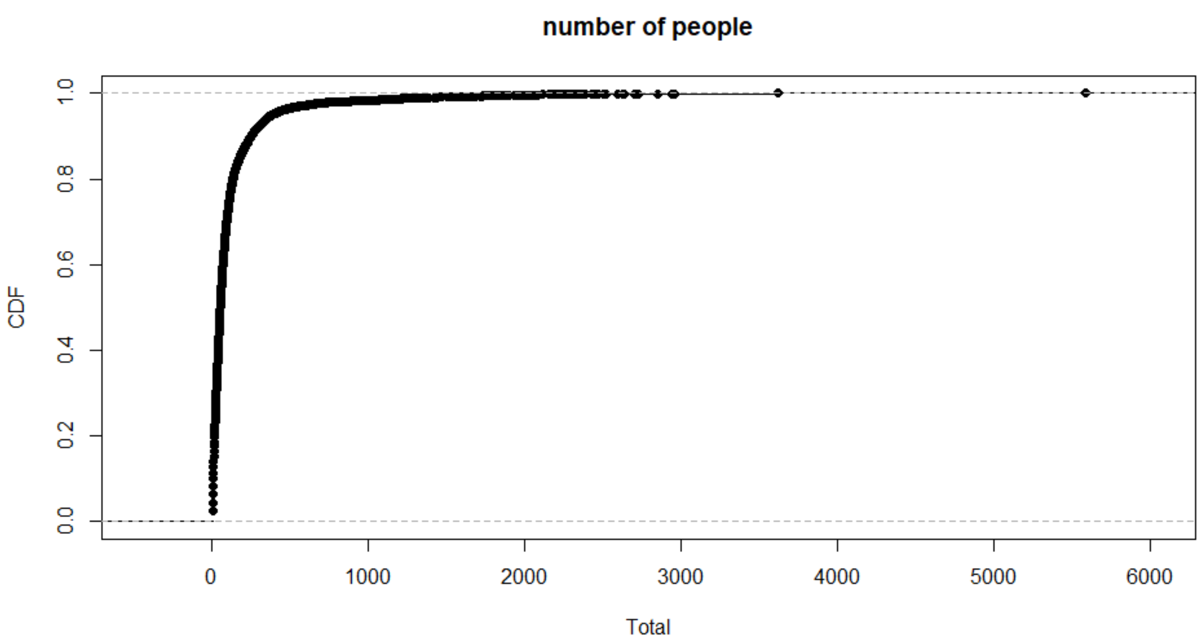


Figure 5.3 Empirical cumulative distribution functions for the response variable

5.4 Methodology

5.4.1 Clustering Analysis Framework

The overall process that we will follow when developing an unsupervised learning model such as GMM can be summarized as shown in Fig 5.4.

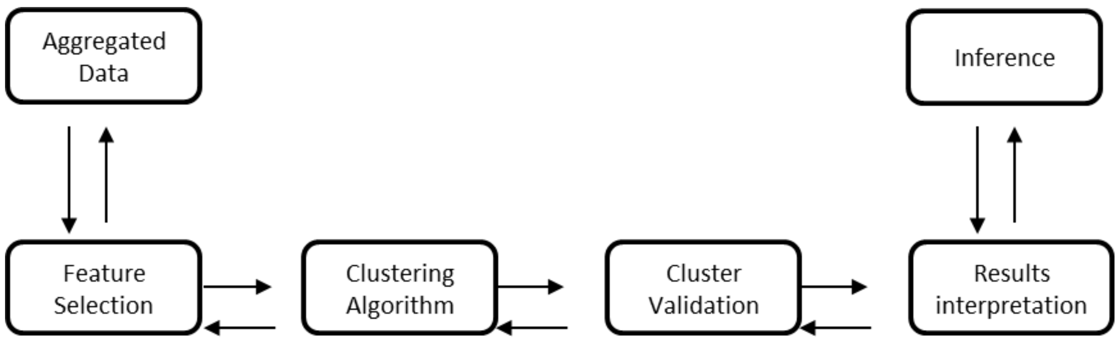


Figure 5.4 Cluster Analysis Framework

Aggregated data involves the data compilation of both the GCFB dataset and the USDA income-related dataset. The distance between each census tract and the assigned food agency and the distance between each family to the assigned food agency will be calculated and saved a variable. After this step, clustering using GMM method is done based on their distances and the demographics mentioned as variables. The clusters obtained are studied and observed.

5.5 Methods

5.5.1 GMM clustering algorithm for comprehensive data characterization

The GMM is a common soft clustering method that can approximate any probability distribution by training several weighted variations of Gaussian distributions and thus increasing the number of mixture components. Each gaussian model in our analysis can be thought of as a coverage class. Here, $Y = [\, Y_1, Y_2, Y_3, \dots, Y_d]^T$, is denoted as an observation vector where, Y_1 is taken as a particular attribute in the given aggregated dataset, and the others are the driver factors. So, d is the number of observation vectors (Ma et al. 2014). The description of GMM is given as follows:

$$p(Y|\theta) = \sum_{k=1}^{K} \alpha_k \phi_k \left(Y|\mu_k \sum_k \right)$$

$$\phi_k \left(Y \middle| \mu_k, \text{å}_k \right) = (2\pi)^{-\left(\frac{d}{2}\right)} \left| \text{å}_k \right|^{-\left(\frac{1}{2}\right)} \exp\left\{ -\frac{1}{2} (Y - \mu_k)^T \text{å}_k^{-1} (Y - \mu_k) \right\}$$

Where k is the number of mixture models, α_k is the mixture weight with $0 < \alpha_k < 1$, and $\sum_{k=1}^{K} \alpha_k = 1$, $\phi_k(y_i|\mu_k, \text{å}_k)$ is the Gaussian model of the kth mixture component, μ_k and å_k denote the mean and the covariance matrix respectively.

Table 5.4 The geometric characteristics of the basic Gaussian models

Model	Distribution	Volume	Shape	Orientation
EII	Spherical	Equal	Equal	-
VII	Spherical	Variable	Equal	-
EEI	Diagonal	Equal	Equal	Coordinate axes
VEI	Diagonal	Variable	Equal	Coordinate axes
EVI	Diagonal	Equal	Variable	Coordinate axes
VVI	Diagonal	Variable	Variable	Coordinate axes
EEE	Ellipsoidal	Equal	Equal	Equal
EVE	Ellipsoidal	Equal	Variable	Equal
VEE	Ellipsoidal	Variable	Equal	Equal
VVE	Ellipsoidal	Variable	Variable	Equal
EEV	Ellipsoidal	Equal	Equal	Variable
VEV	Ellipsoidal	Variable	Equal	Variable
EVV	Ellipsoidal	Equal	Variable	Variable
VVV	Ellipsoidal	Variable	Variable	Variable

Each Gaussian model represents a cluster, and the 14 models proposed (Fraley & Raftery, 2007; "Mathematical Statistics and Data Analysis - John A. Rice - Google Books", n.d.) are shown in table 5.4. Hence the parameter set of a GMM is composed of $\{\alpha_k, \mu_k, \Sigma_k\}$, with $1 \leq k \leq K$. The parameters are estimated in the maximum likelihood setting. The optimization is usually carried out using the Expectation-Maximization (EM) algorithms, which are depicted as follows in two steps as follows:

Expectation Step: from the below equation, a posteriori probability γ_{jk} at the jth data value ($j \in [1, N]$, N denotes the number of samples) is computed based on the randomly given initial values of $\{\alpha_k, \mu_k, \Sigma_k\}$:

$$\gamma_{jk} = \frac{\alpha_k \phi_k(X|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \alpha_k \phi_k(X|\mu_k, \Sigma_k)}$$

Maximization Step: In this stage, new set of values for the parameters $\{\alpha_k, \mu_k, \Sigma_k\}$ can be obtained with γ_{jk} designed from the earlier mentioned Expectation Step as:

$$\alpha_k = \frac{\phi_k}{N}$$

$$\mu_k = \frac{1}{\phi_k}\sum_{j=1}^{N} \gamma_{jk} X_n$$

$$\sum_k = \frac{1}{\phi_k}\sum_{j=1}^{N} \gamma_{jk}(X_n - \mu_k)(X_n - \mu_k)^T$$

Where $\phi_k = \sum_{j=1}^{N} \gamma_{jk}$. A fresh set of $\{\alpha_k, \mu_k, \Sigma_k\}$ can be obtained by Maximization Step, which is applied to the earlier Expectation Step to obtain the new γ_{jk}. Both these steps are hence iteratively calculated until convergence is obtained based on the likelihood function:

$$L(\theta|X) = \sum_{i=1}^{N} ln\left\{\sum_{k=1}^{K} \alpha_k \phi_k \left(Y|\mu_k \sum_k \right)\right\}$$

Under the pre-set K, the final dataset of $\{\alpha_k, \mu_k, \Sigma_k\}$ is calculated by the maximum $L(\theta|X)$, and each $\phi_k\ (Y|\mu_k \sum_k)$ is termed a cluster. To obtain the optimal number of clusters (K), the following methods are considered.

5.5.2 Optimal number of Clusters

A key role in the GMM clustering method is selecting the standard Gaussian model (table 2) from the 14 types of basic models proposed. One model is chosen at each clustering point, and the data distribution is explained. The best basic model helps to achieve the best clustering performance. Since two or more variables may have a positive or negative correlation, the orientation of the covariances is constrained to be variable across classes. As a result, the EEV, EEE, EVI, and EII models were chosen for the GMM clustering analysis.

Another important function in the GMM clustering approach is to determine the optimal number of components (K), which can be obtained through two techniques – The Bayesian information criterion (BIC) (Srivastav, Tewari, and Dong 2013) and the Silhouette score (Rousseeuw, 1987). For BIC, the criterion is formulated as follows:

$$BIC = L(\theta|X) - \frac{M}{2}\log(N)$$

Where N is the number of samples. The total number of free parameters is represented by M and this is obtained as below:

$$M = Kd + \frac{1}{2}Kd\,(d+1) + (K-1)$$

Where d is the number of observation vectors. The K value and the most suitable model that maximizes the BIC typically represents the optimal (K) for the model. Fig 5.5. Provides the results of the BIC values obtained for the chosen gaussian models.

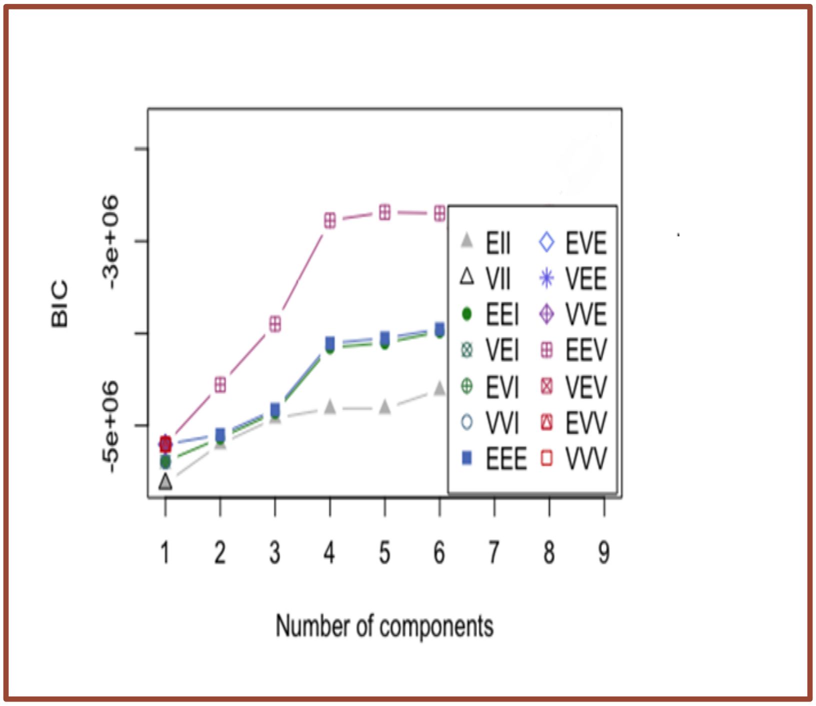


Figure 5.5 Plot of BIC values for a variety of models and a range of number of clusters

Conferring to [9], the silhouette score can also measure the goodness of any clustering technique. In the silhouette score, there is a term called $s(i)$ which is calculated as follows :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Where $a(i)$ is the average dissimilarity of the data value i (can be any variable in the dataset) with all the other data within the same cluster. $b(i)$ is the lowest average dissimilarity of i to any other cluster. Fig 5.6. Shows the average silhouette score for different number of clusters.

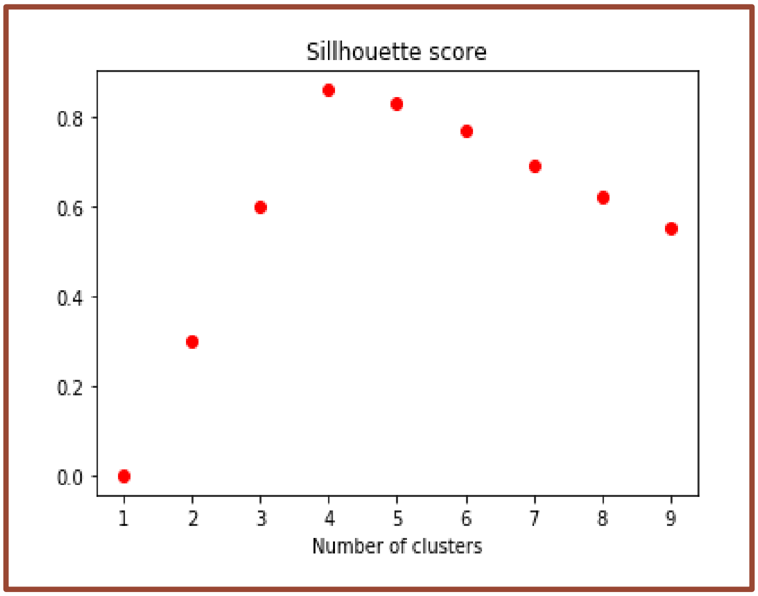


Figure 5.6 Average silhouette scores for different number of clusters

From Fig 5.6. we observe that the average silhouette score is the highest when the number of clusters is 4 and in Fig 5.5, we observe that after a steep rise from clusters 3 to 4 it has been relatively steady when the number of clusters recorded 4. Hence, from BIC, model EEV (Ellipsoidal, equal volume, and equal shape) with 4 clusters is taken as best blend.

5.6 Predictive Modelling Framework

Figure 5.7 depicts the study's framework for predictive modeling. The input data preparation is the first step in the study; during this phase, we will add an additional input variable to the food demand dataset consisting of the clustering results from the previous clustering analysis study. We will investigate the effect of GMM clustering findings on the predictive accuracy of the food demand dataset in this way. As a result, predictive modeling is performed on two datasets: one containing the clustering results as an additional input variable, and another containing the clustering results but not the clustering results.

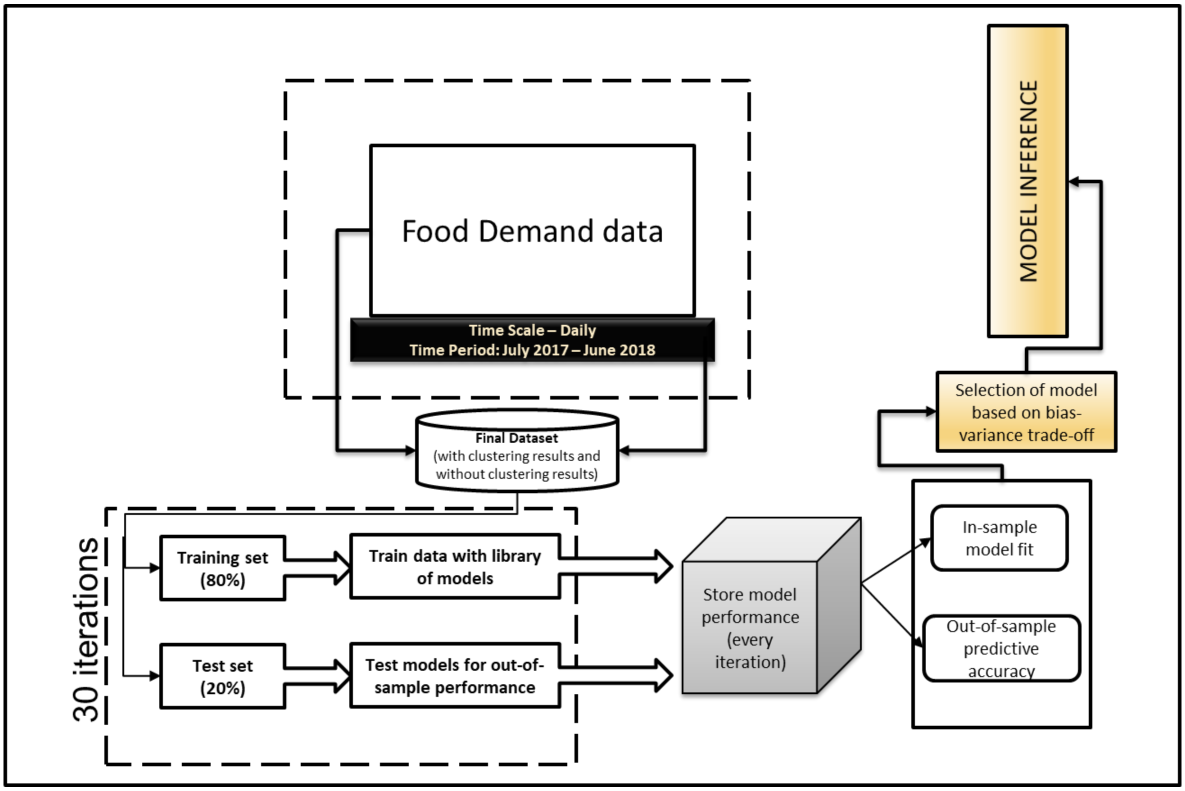


Figure 5.7 Predictive Modeling Framework

As evident from this Fig 5.7, while data specific to the GCFB was used to demonstrate the applicability of the proposed research, the approach and methodology is transferable and can be extended to other food banks and regions.

5.7 Prediction Models

Numerous types of parametric, semi-parametric and non-parametric machine learning methods have been applied and trained to the food demand dataset (both with and without clustering results). This is done to develop optimum predictive models that portray the best understanding of the complex and non-linear relationships between the demand of food in the food banks and the various input variables. We utilized the methods of Generalized Linear Model (GLM), Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and Bayesian Additive Regression Trees (BART) to estimate the food demand in the given region that GCFB handles. Based on these machine learning algorithms, predictive models of the food demand are developed employing rigorous cross-validation to highlight the model that out-performed all the others in terms of out-of-sample predictive accuracy. A brief review of each of the methods used in our study are examined below.

5.7.1 Generalized Linear Model (GLM)

GLM stands for Generalized Linear Models and is an extension of linear regression. The normality assumption is relaxed in GLMs, enabling the response variable to be distributed according to an exponential family of distributions (e.g., Gaussian, Binomial, Poisson, Gamma, or inverse-Gaussian) and linked to the predictors through a link function (Cordeiro & Mccullagh, 1991; McCulloch, 2000). A dependent variable Y with a distribution that falls into the categories of normal, binomial, Poisson, gamma, or Inverse-Gaussian, as shown in the equations below:

$$Y_i \sim f_{Y_i}(y_i)$$

$$f_{Y_i}(y_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right\}$$

where θ and ϕ are the location and scale parameters respectively.

A set of independent variables x_i.

A link function *g(.)* binding the parameters of the dependent variable to the linear combination of input variables.

GLMs are widely popular due to their ease of usage and interpretability. However, GLMs consider the ‘rigid’ assumptions of global parametric models by this means costing the performance of predictive accuracy.

5.7.2 Generalized Additive Model (GAM)

Generalized Additive Model (GAM) is a semi-parametric machine learning method. It relaxed the assumption of linearity that is considered in the above mentioned GLM method, thereby allocating for regional non-linearities (Hastie & Tibshirani, 1990; Hastie & Tibshirani, 1986). Here, the dependent variable y has a distribution with mean $\mu = \mathrm{E}[Y|x_1, x_2, \ldots\ldots x_p]$ (an assumption GAMs makes) associated to the predictor variables via a link function as:

$$g(\mu_i) = \alpha + \sum_{j=1}^{p} f_j\,(x_j)$$

where each f_j is a smoothing function of a class of functions projected non-parametrically, like regression splines and tensor product splines.

5.7.3 Multivariate Adaptive Regression Splines (MARS)

MARS is a semi-parametric, adaptive, and compliant regression technique that is well suited for high-dimensional problems (i.e., datasets with a large number of input variables) (Friedman, 1991). It can be thought of as a stepwise linear regression simplified. As shown in the equation below, the MARS model uses sum-of-splines to allow the answer to vary non-linearly with the input variables:

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m\,(X)$$

where each $h_m(x)$ represents the reflected pair of linear splines, β_0 represents the intercept and β_m represents the vector of the coefficients. β_m coefficients are projected by reducing the sum of square errors. MARS is built in a forward manner implementing cross-validation to choose the optimal collaboration of variables and avoiding over-fitting.

5.7.4 Random Forest (RF)

RF is a non-parametric tree-based ensemble data-miner (Breiman, 2001). It is a modification of bootstrap aggregation to multiple Classification and Regression Trees (CART) and taking the mean of the predictions of the roughly uncorrelated trees to produce the final estimation. The procedure consists of B bootstrapped regression trees (T_b) with B chosen based on cross-validation. Regression trees are low bias high variance techniques. In other words, they can obtain the shape of the data pretty well (low bias) but are highly vulnerable to outliers (high variance) (Hastie et al., 2003). RF pulls in model averaging as a variance reduction technique. The final estimate is, therefore, the average of predictions across all B trees as shown below.

$$f_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

In RF, each tree is developed using bootstrapped re-samples of the input data, and split variables are also randomly chosen to encourage independence between the trees. Hence, RF can achieve strong predictions by lowering the correlation between the trees such that model averaging can be used to get a low-bias, low-variance predictions, and keeping the errors of every individual unpruned tree low.

5.7.5 Bayesian Additive Regression Trees (BART)

Bayesian Additive Regression Trees (BART) is a non-parametric Bayesian method. The BART model implements sum-of-trees to estimate the dependent variable. Each regression tree is created by separating the data area recursively into sub-regions or nodes and tailoring a simple model (e.g., mean of the response or dependent variable) in each region. The covariates that are decided to split, and the split values chosen is done such that the best fit is achieved at each sub-region. The final model estimate contains the summation of the estimate from m small trees, as shown in the equation (Merwe, 2009):

$$Y = (\sum_{i=1}^{m} g(x; T_j, M_j)) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

where $g(x; T, M)$ is the function which designates the parameters of the terminal nodes of tree T, $\mu_i \in M$ to the predictors x. To ensure that each tree contributes only partially to the final predictions, regularization priors are used to control the model's complexity. Regularization priors, therefore, help eliminate an individual tree's effect being excessively influential on the sum-of-trees model (Merwe, 2009). BART is robust to outliers, has good predictive power, and is a completely probabilistic approach which implies it can yield complete distributions of the predicted response values.

5.7.6 Bias-variance tradeoff

The ability of a predictive model to make good predictions on an individual test sample determines its generalization efficiency. The biggest decision maker for minimizing generalization error is balancing the bias-variance trade-off (Hastie et al., 2003). One of the most commonly used approaches for matching bias and variance is cross validation. To approximate predictive precision, we use the k-fold cross validation technique. K-fold cross-validation involves slicing the data into *k* equal-sized subsets at random. The model is fitted on all subsets except the *k*th held-out sample of each replication, and the predictive accuracy is determined based on the performance of the models on the kth held-out subset. The efficiency of the out-of-sample model was calculated in this paper using a 30% holdout cross validation and the formulae below:

$$MSE_{out-of-sample} = \frac{1}{k}[\sum_{k=1}^{n}\frac{1}{m}(\sum_{i=1}^{m}(y_{i,k} - \widehat{y_{i,k}})^2)]$$

$$MAE_{out-of-sample} = \frac{1}{k}[\sum_{k=1}^{n}\frac{1}{m}|\sum_{i=1}^{m}(y_{i,k} - \widehat{y_{i,k}})|]$$

k = number of times cross-validation was done

m = hold-out numbers during each cross-validation

$y_{i,k}$ = during the *kth* cross-validation, the *ith* actual observation that was kept out at random

$\widehat{y_{i,k}}$ = using the model developed, using the training set data during the *kth* cross-validation, and obtaining the predicted *ith* observation

In this paper, we pick models based on both in-sample fit and out-of-sample prediction accuracy. The in-sample MSE (Mean Square Error), MAE (Mean Absolute Error), and adjusted R^2 were used to calculate the in-sample error, while the out-of-sample MSE (Mean Square Error) and MAE (Mean Absolute Error) were calculated as previously stated.

5.7.7 Tuning parameters

Generalized Linear Model (GLM): For GLM, the value of the tuning parameters applied are listed below:

***k*:** Refers to the number of degrees of freedom used for the penalty. For our study, when k = 2 the best Akaike Information Criterion (AIC) value is obtained: k = log(n). Hence, in our study we used k=2.

Dist.= Gaussian: This parameter specifies the type of error distribution and link function to be used in the model. In this study, for all the different methods stated, we assumed that the error follows "Gaussian distribution" and the link function is taken to be an "identity" function.

Generalized Additive Models (GAM): We implemented a stepwise update methodology and cubic smoothing function which generated the best predictive accuracy among the tuning options to select the best fit model.

Multivariate Adaptive Regression Splines (MARS): The tuning parameters used for developing the MARS model are described below:

$\boldsymbol{pMethod}$: It refers to the pruning method. The type of pruning method used is "cv". $pMethod = cv$ uses cross-validation to select the number of terms. This selects the number of terms that gives the maximum mean out-of-fold R^2 on the fold models. We selected the model based on the best goodness-of-fit for the models.

$\boldsymbol{nfold}$: This parameter refers to the number of cross-validation folds. In R, default is 0 i.e., no cross validation. If $nfold > 1$, earth first builds a standard model as usual with all the data; then it builds $nfold$ cross-validated models, measuring R^2 on the out-of-fold (left out) data each time. The final cross validation R^2 $(CVRSq)$ is the mean of these out-of-fold R^2. The above process of building $nfold$ models is repeated $ncross$ times. In our research, we used the number of cross-validation folds as 10.

$\boldsymbol{ncross}$: This parameter only applies if $nfold > 1$. Each cross-validation has $nfold$ folds. The default in R is 1 and in our research we used the value as 5.

Random Forest (RF): The tuning parameters for the RF model are considered below.

$\boldsymbol{mtry}$**:** The number of variables randomly sampled as candidates at each split while growing the trees. To be noted that the default values for the regression tree is $p/3$, where p is the number of independent variables used in the model.

$\boldsymbol{ntree}$: This parameter refers to the number of trees to grow. This must not be set to a very small number to guarantee that every input row will get predicted at least a few rounds. In our research, we selected the value of the $ntree$ that yielded the least mean square error (mse) while growing the trees.

Bayesian Additive Regression Trees (BART): The tuning parameters used in the BART model are described below:

$\boldsymbol{k}$**:** For regression, k determines the prior probability that $E(Y|X)$ is contained in the interval $(y\_\{min\}, y\_\{max\})$, based on a normal distribution. For example, when we have $k = 2$, we get the prior probability to be 95%. For classification, k determines the prior $E(Y|X)$ between $(-3,3)$. Note that a larger value of k results in more shrinkage and a more conservative fit.

$\boldsymbol{nu}$**:** It refers to the degrees of freedom for the inverse χ^2 prior.

$\boldsymbol{q}$**:** This parameter refers to the quantile of the prior on the error variance at which the data-based estimate is assigned. It is to be stated that greater the value of q, the more forceful is the fit; this is because it corresponds to placing more prior weight on values lower than the data-based estimation . It is not used for classification.

$\boldsymbol{m}$**:** This parameter refers to the number of trees to be grown in the sum-of-trees model.

5.8 Results and Discussion

5.8.1 The results of GMM clustering

The GMM parameterized by EM algorithm is applied to the aggregated dataset containing various socio-economic and demographics details of the region supported by GCFB. As shown in Table 4. The data is divided into 4 clusters, the four clusters have been named based on the proximity of the families with their respective food agencies. It is seen that the average distance between cluster 1 and 2 is 1.03 miles, between 2 and 3 is 3.11 miles and between 3 and 4 is 14.65 miles. These clusters are reasonable considering that the dataset is intended for finding factors concerning food insecurity for an urban area.

5.8.2 Summarization of clustering patterns

In Table 5.5, we observe that the further away the families are from the food agencies, the more people there are in the family. This holds true for children and especially adults, where as not so much for the seniors. From this table, we can interpret the number of children and seniors having low access toward food resources. It is seen that around 13,967 children live very far away from the food agencies while 122884 children live less than 0.42 miles away from the nearest food assistance. Also, the number of tracts increases as the distances from the food assistance increases. This makes sense since they are more scattered over the urban areas. Regarding the income of families in each cluster, we can see that most families (89.1%) live in tracts that are considered poor (tracts are considered poor if their average household income is less than Ohio's median income, according to (Mattogno et al., 2014). It is also seen that around 11% of the families are living in tracts having high income levels.

Table 5.5 Detailed calculation results of GMM clustering

Variables	**Cluster 1 - proximate**	**Cluster 2 – reasonable distance**	**Cluster 3 - Distant**	**Cluster 4 – extremely distant**	**Total**
Number of families	197,844	199,475	195,081	17,009	609,409
Number of adults	221,523	244,783	245,107	23,431	734,844
Number of children	122,884	154,202	152,667	13,967	443,720
Number of seniors	120,513	127,827	130,792	10,417	389,549
Number of people	464,920	526,812	528,566	47,815	1,568,113
Average number of adults in family	1.12	1.23	1.26	1.38	1.21
Average number of children in family	0.62	0.77	0.78	0.82	0.73
Average number of seniors in family	0.61	0.64	0.67	0.61	0.64
Average number of people in family	2.35	2.64	2.71	2.81	2.57
Number of tracts	464	545	654	884	943
Average Distance (miles)	0.42	1.45	4.63	19.47	2.64
Coverage	32.5%	32.7%	32.0%	2.8%	100.0%
Pct of Poor People	92.8%	91.2%	84.3%	74.7%	89.1%
Pct of Rich People	7.2%	8.8%	15.7%	25.2%	10.9%

The clustering has contributed to the discovery that the farther families are from food, the more likely they are to live in a tract with a high-income population. The coverage rate is the number of families within a mile that are served by at least one food bank. The research by Alotaik et al. (2017) is the only one we are aware of that did a similar analysis. However, they introduced census tracts, which are more common than the family's individual locations. It can be seen that to increase the coverage of supply of food assistance to the poor people located very far away, some of the food agencies should be moved from cluster 1 to cluster 4 where there are areas with less coverage and people with low income. The specific tracts and locations of families are not presented considering the huge amount of data provided for this region, but they are very well known and can be easily identified in the given dataset.

Fig. 5.8. shows the spread of the agencies and families in each cluster. The latitude and longitude information were used to plot the corresponding graph for each cluster on the map of the Ohio region. It is clearly visible that the distance increases in each cluster with the average distances of each cluster mentioned in Table 5.5.

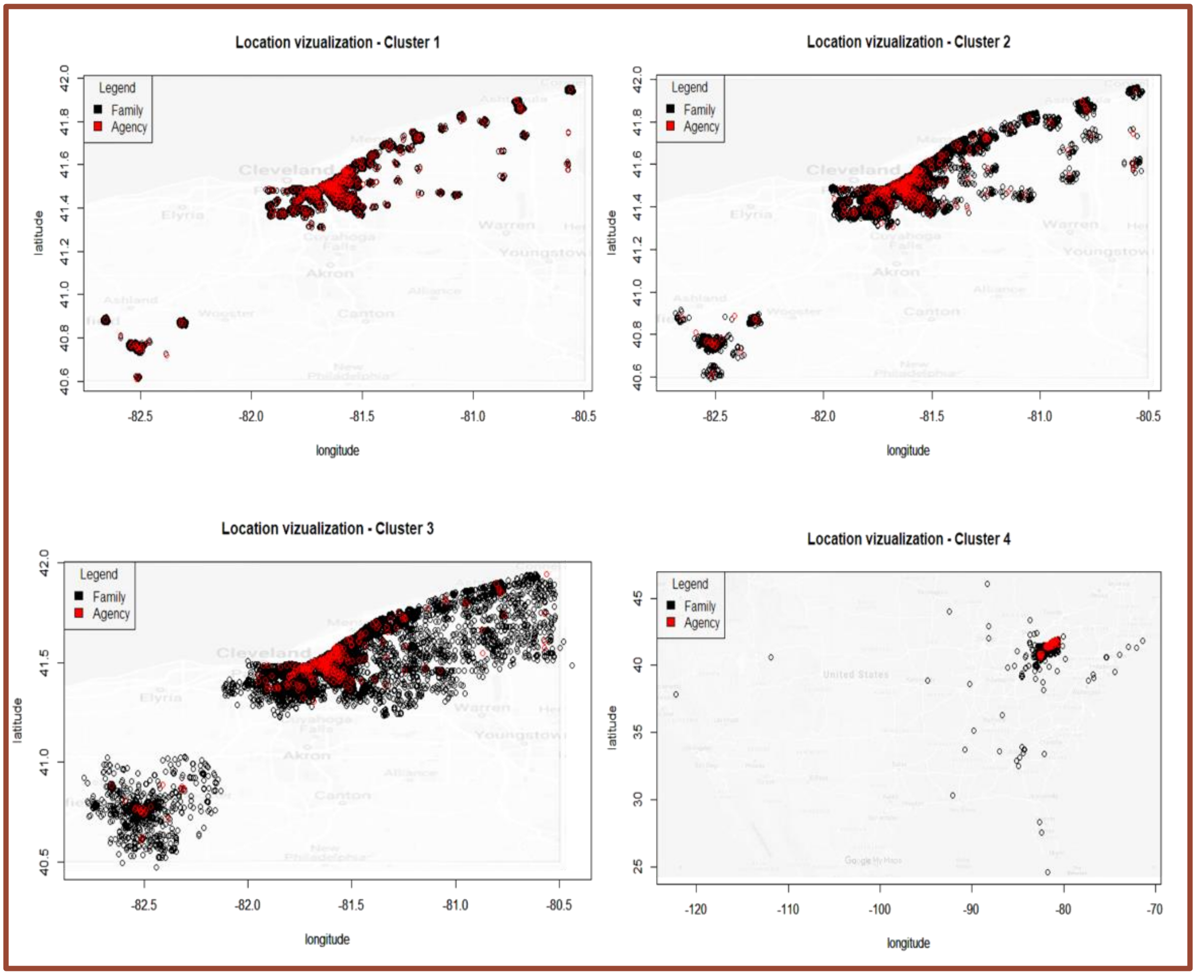


Figure 5.8 Distance between families and agencies in each cluster

5.8.3 Modeling of donated food demand

We developed predictive models for the food demand study considering clustering results information as another independent variable in the food demand dataset (Table 5.2.) and without clustering results to observe if there is any accuracy improvement in the prediction results. We trained the food demand dataset (with and without clustering results separately) with the methods of Generalized Linear Model (GLM), Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and Bayesian Additive Regression Trees (BART). In this section we will discuss the performance of each of the trained models and choose the one with the final model based on the one that has the best out-of-sample predictive accuracy. Table 5.5 and 5.6 summarizes the goodness-of-fit, and predictive performance of each of the trained models. The percentage improvement yielded by each of the trained models over having no statistical model and using the historical average as a predictor (i.e., the ‘mean-only’ model) is provided in Table 5.7.

It can be seen that BART substantially out-performs all the other models in terms of goodness-of-fit. Comparing our results of the GLM with the results of BART supports our hypothesis that linear models don’t capture the complex non-linearities in food demand data adequately.

The plots of predicted versus observed food demands are given in Fig 5.9 for the data included with clustering results and Fig 5.10 and for the data excluding the clustering results. In the case of the former, the 95 % credible intervals provide 57.31% coverage for all the observations whereas the 95% prediction interval offers a 97.68% coverage (Fig 5.9.). In case of the data without clustering information, the 95% credible intervals provide 20.65% coverage for all the observations whereas the 95% prediction interval offers a 95.98% coverage (Fig 5.10.).

By observing the results, it can be concluded that although BART does provide the best predictive accuracy for both datasets, the dataset without clustering results has an unsatisfactory overall error level. As can be seen in the fig 5.9, the results of the models for dataset consisting of clustering results have been greatly improved. However, the models tend to underestimate the more extreme ends of demand.

Table 5.6 Modelling with clustering results

Model	**Tuning Parameters**	R^2	In-sample		Out-of-sample	
			RMSE	MAE	RMSE	MAE
Mean (Null model)	-	-	-	-	443.12	328.37
Generalized Linear Model (GLM)	k=2.0, Dist.=Gaussian	0.55	233.65	189.82	247.21	193.92
Generalized Additive Model (GAM)	Stepwise update	0.58	199.23	174.63	213.51	183.26
Multivariate Adaptive Regression Splines (MARS)	pMethod: cv; nfold: 10; ncross=5	0.42	217.72	195.27	248.56	199.33
Random Forest (RF)	mtry=p/3 =3; ntree=100	0.63	142.82	134.19	155.94	140.32
Bayesian Additive Regression Trees (BART)	**k=2,nu=10,q=0.75,m=50**	**0.82**	**137.17**	**99.31**	**143.57**	**105.11**

Table 5.7 Modelling without clustering results

Model	**Tuning Parameters**	R^2	In-sample		Out-of-sample	
			RMSE	MAE	RMSE	MAE
Mean (Null model)	-	-	-	-	683.34	510.39
Generalized Linear Model (GLM)	k=2.0, Dist.=Gaussian	0.28	379.36	354.49	399.53	357.51
Generalized Additive Model (GAM)	Stepwise update	0.31	363.69	335.62	378.57	329.26
Multivariate Adaptive Regression Splines (MARS)	pMethod: cv; nfold: 10; ncross=5	0.35	312.23	289.60	340.41	299.46
Random Forest (RF)	mtry=p/3 =3; ntree=100	0.41	278.33	241.91	281.26	263.51
Bayesian Additive Regression Trees (BART)	**k=2,nu=10,q=0.75,m=50**	**0.47**	**226.61**	**202.11**	**235.36**	**216.44**

Table 5.8 Percentage improvement over the ‘null’ model for modelling with clustering results

Models	Out-of-sample error (%imp)	
	RMSE	MAE
GLM	44.2	40.9
GAM	51.8	44.2
MARS	43.9	39.3
RF	64.8	57.2
BART	**67.6**	**67.9**

Table 5.9 Percentage improvement over the ‘null’ model for modelling without clustering results

Models	Out-of-sample error (%imp)	
	RMSE	MAE
GLM	41.5	29.9
GAM	44.6	35.4
MARS	50.2	41.3
RF	58.8	48.3
BART	**65.5**	**57.5**

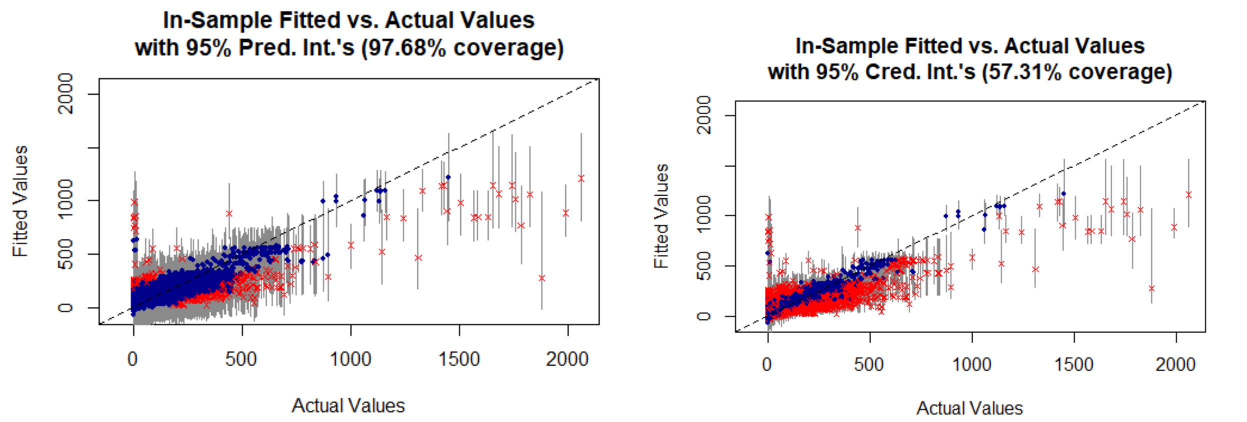


Figure 5.9 The prediction results of food demand dataset with clustering results

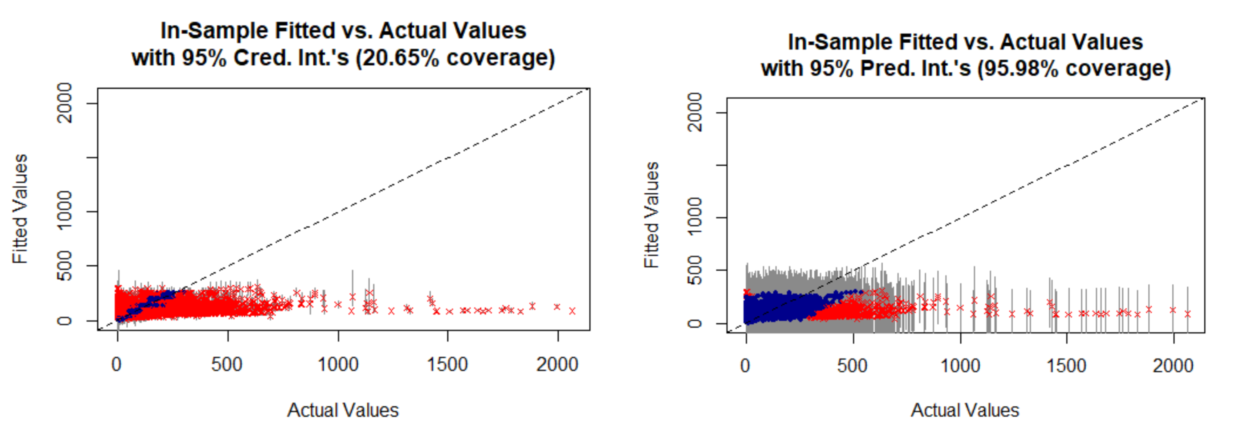


Figure 5.10 The prediction results of food demand dataset without clustering results

5.9 Discussion

On observing the results in Section 5.5, we see how clustering and the usage of clustering results in a predictive modelling of food demand aids in the better forecasting accuracy. This leads us to implement a research framework that is streamlined to be used by the food bank officials. The research framework proposed is termed as the two-stage hybrid demand estimation model to identify and classify the aggregated dataset to clusters and using the cluster results on the food demand dataset (obtained from aggregated dataset) for predictive modelling by Bayesian Additive Regression Trees (BART). The outcome of this framework is to understand and aid the food bank management with the food demand behavior with greater accuracy and optimal planning.

Figure 5.11 depicts the proposed approach for developing the data-driven, demand estimation model. The major steps include the data collection and data wrangling followed by the implementation of the algorithm-based statistical learning methods for classification and prediction.

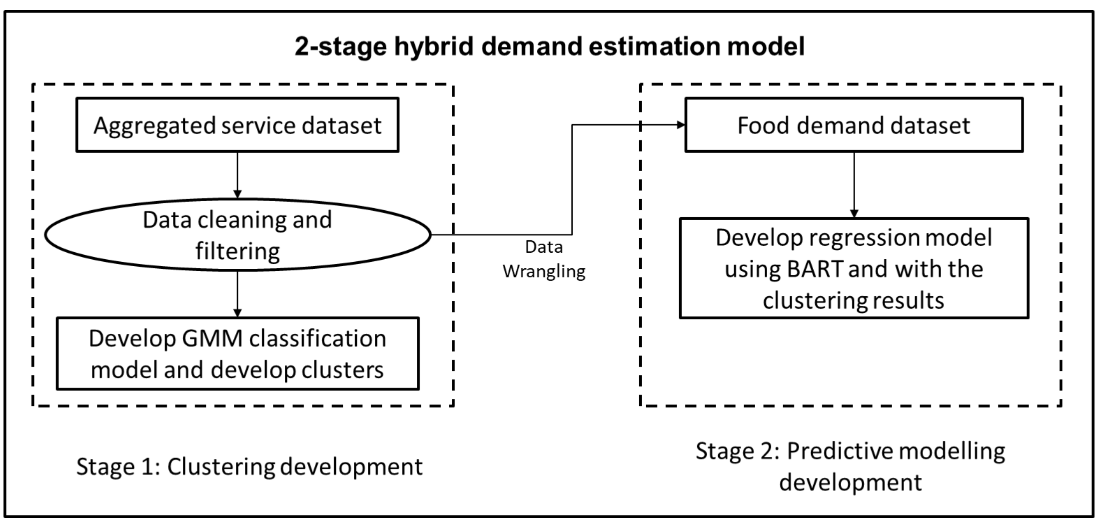


Figure 5.11 Flow chart of research methodology steps for developing two-stage hybrid demand estimation model

5.10 Conclusions

This paper proposes that the characteristics of a particular region by means of a clustering method such as GMM in terms of accessibility of food assistance and finding ways to increase their access to food. With this information, GCFB can manage and distribute their food resources to the food insecure in an efficient and equitable manner by targeting the regions of food assistance desserts, increasing the coverage in regions of people receiving low income and located far away from the source of food assistance. By the results of the food accessibility pattern study, food demand of the GCFB organization is studied by developing predictive models. It is seen that by implementation of clustering results to the predictive models have an obvious accuracy improvement and hence a two-stage hybrid demand estimation model is proposed based on the results obtained. A future direction in this research is the utilization of these predictive model results of food demand in a given region as input parameters to mathematical models developed to improve the equitable distribution of donated foods to the people in need.

6. VISUAL ANALYTICS FOR DECISION-MAKING FOR FOOD BANK SUPPLY CHAIN PLANNING UNDER UNCERTAINTY: CONCEPTUAL FRAMEWORK AND EXPERIMENT

6.1 Introduction

Food insecurity is characterized as a situation in which a household lacks enough food to keep one or more members active and safe. During the year 2014, an estimated 14 percent of households in the United States were food insecure, with 5.6 percent having very poor food security, meaning that their eating habits were disordered due to a lack of money and other sources for purchasing food (Feeding America, 2015). In recent years, 40% of food in the United States has gone uneaten, equating to 20 pounds of food per month per human (Irani et al., 2018). In the United States, the equivalent of $165 billion of food is discarded per year, which ensures that the uneaten food winds up decaying in landfills as the single largest part of urban solid waste, accounting for a significant portion of U.S. methane emissions. About 25% of the food and alcohol purchased by American families were wasted (Hall et al., 2009). When one of every six Americans lacks a stable supply of food, a 15% reduction in food losses will provide enough food to sustain more than 25 million Americans every year (USDA, 2020). Hence, optimizing the efficiency of the U.S food system requires a collaborative, systematic approach by several entities (government, businesses, and consumers).

In the United States, there are a number of initiatives that bring together government, corporate, and private institutions to help citizens who have insufficient access to food. Feeding America is one of the country's leading non-profit hunger reduction organizations, fighting hunger and food shortages. Feeding America has a nationwide network of about 200 food banks and 60,000 food pantries and meal services that provide food and assistance to the hungry (food agencies in the supply chain). Donations to food banks come from a multitude of places, including big food providers and retailers. Using leased or owned trailers, the donated food is delivered back to the food bank. These donated foods are stored at the food bank to ensure that they are of high quality. Following that, the trucks are used to deliver the quality-inspected foods to the food agencies depending on their accessibility. Food agencies provide the public with donated goods. Food banks, in general, act as wholesalers of surplus food. They collect bulk contributions from society, government, and private supporters, store and warehouse items, and then distribute to food banks. Since the situation is based on donations, the supply chain is complex because there is a problem matching supply to demand. The availability, for example, is highly unpredictable without understanding the frequency, quality, and amount of donated goods in advance. Similarly, without taking into account the conditions that contribute to hunger (poverty, distance, etc.), the need for food is extremely unpredictable. As a result, the purpose of a donation-driven supply chain like the food bank supply chain is to increase relief while minimizing food waste. Using leased or owned trailers, the donated food is delivered back to the food bank. These donated foods are stored at the food bank to ensure that they are of high quality. Following that, the trucks are used to deliver the quality-inspected foods to the food agencies depending on their accessibility. Food agencies provide the public with donated goods. Food banks, in general, act as wholesalers of surplus food. They collect bulk contributions from society, government, and private supporters, store and warehouse items, and then distribute to food banks. Since the situation is based on donations, the supply chain is complex because there is a problem matching supply to demand. The availability, for example, is highly unpredictable without understanding the frequency, quality, and amount of donated goods in advance. Similarly, without taking into account the conditions that contribute to hunger (poverty, distance, etc.), the need for food is extremely unpredictable. As a result, the purpose of a donation-driven supply chain like the food bank supply chain is to increase relief while minimizing food waste.

We bridge this gap by using visual analytics and developing successful data visualization as a tool in food bank officials' decision-making processes. Visualizing and interpreting data is critical for obtaining valuable knowledge that isn't accessible through quantitative analysis. Visualization tools depict complex patterns that cannot be represented in any other way, making them an important tool in the decision-making process (Delpish & Jiang, 2019). It assists decision makers in quickly analyzing large volumes of data, quickly identifying patterns and problems, sharing ideas with key players, and influencing decisions.

In recent times, non-profit supply chains such as food banks, etc., have never been so complex and disturbed by hazards and variabilities (after-effects of COVID-19 pandemic such as rise of unemployment, increased health hazards, etc.) (Providence, 2020). Additionally, most of the decision support systems (DSS) available in the literature do not manage the issues of uncertainty and assume restrictive hypotheses (Beamon & Balcik, 2008). To develop a systematic, sustainable infrastructure that promotes co-ordination at various levels of planning, visual management is considered. Dashboards are an example of visual management in action, as they proactively promote and enhance decisions, keep decision-makers focused on the most important problems at their level of preparation, and assist them in improving efficiency and achieving desired results (Stirrup, 2014). The main information presented on dashboards is expected to have an impact on every organization's decision-making activities, so the dashboard's design is critical. The word dashboard is derived from the dashboard of a car, which displays the metrics that the driver needs to know; similarly, dashboards display data that allows managers and employees to visually recognize supply and demand trends, patterns, and anomalies. There has been research done on the evolution, functions, categories, and formats of dashboards, as well as the relevance of dashboard design and production (Ko & Chang, 2018; Delpish & Jiang, 2019; Raffensperger et al., 2020; Wu et al., 2020). However, the development process should be explored, considering potential issues like data availability and reliability, that would lead to some process change or improvement and should be considered as part of the dashboard development initiative.

The current literature focused on the development of dashboards is too complex for non-profit organizations. In most cases, Food banks do not have, an organized management system, manifesting a low level of statistical knowledge and maturity among the food bank employees (Desai, 2015). This is reflected in several aspects, including the level of maturity of information systems, compromising visual management. Desai (2015) listed other important food bank management characteristics in a literature review, including the following, which may have a greater effect on visual management: Non-profit organizations, such as food banks, are limited in terms of financial and human resources; decision-making is more intuitive and based on experience; and most operations are regulated by informal rules and procedures with little standardization and formalization. As a result, this paper provides a systematic-based procedure for developing dashboards for food bank organizations, with the goal of enhancing food bank administrators' decision-making on food and resource allocations. The procedure is an adaptation of the traditional product development process by Pahl and Beitz (2013).

The paper is structured as follows: in Section 6.2 a brief review of the literature is presented, Section 6.3 presents the dashboard development procedure; Section 6.4 presents the implementation of the proposed procedure in a food bank organization. Finally, in Section 6.5 the conclusions and suggestions for future work are presented.

6.2 Literature Review

There is considerable relevant literature discussing the role of machine learning and data visualization models in the improvement of uncertainty of several aspects of the supply chain. However, hunger relief operations receive scant coverage. The reviewed literature reveals that supply and demand traceability, as well as visualization of food insecurity, are key performance indicators (KPIs) that come under the category of reducing uncertainty (Shi et al., 2020). In the food relief operations of the Food Bank of Central and Eastern North Carolina, Davis et al. (2016) used time series estimation techniques, moving average, and exponential smoothing to estimate the amount of food donated per definition of food per category of donor. According to their data analysis, exponential smoothing method had provided better prediction results than the other established methods to predict the food donation. Phillips et al. (2013) introduced an empirical model to predict the total quantity of food obtained by a food bank in north central Colorado. The presented model is a threshold model in which the Generalized Pareto distribution is used, and the food donated by food suppliers is modeled based on the donors' characteristics. The main focus of this research was on identifying the demand-supply gap and possible solutions. Using Markov Chain analysis, Jiang et al. (2013) investigated various data mining techniques to analyze the donation pattern and stochasticity. Brock and Davis (2015) conducted another analysis on the predictive modeling of donations received from supermarkets, focusing solely on supermarket transactions and using both conventional and nontraditional forecasting approaches. Recently, (Nair et al., 2017; Brock & Davis, 2015; Nuamah, 2016) evaluated several different approximation methods to estimate the daily availability of food based on a set category of foods and food providers and considering only the correlation between food types donated. (Alotaik et al., 2017; Sucharitha & Lee, 2019) developed clustering algorithms to better understand the possible causes of food insecurity in a given region by means of understanding the characteristics and structure of the food assistance network in a particular region. However, despite the fact that these methods and studies were useful, they were not presented as a decision-making tool for effective decision-making. Effective data visualization and visual analytics will help you prepare effectively. The ability to visualize and analyze data is crucial for gaining valuable information that goes beyond quantitative analysis. Visualization software depicts complex patterns that cannot be represented in any other way.

Desai et al. (2015) developed interactive dashboards to enable judicious decision making for optimal food bank operations to meet hunger needs. Visual analytics was further extended by Hindle and Vidgen (2018) where an analytics methodology was provided to develop a dynamics visualization tool. However, both these research methodologies do not take into consideration the various levels of planning in their dashboard development and have focused on the strategic level of planning meant for the higher management only. According to the Salesforce Nonprofit Trends Survey, more than half of nonprofits (53 percent) find it simple to collect program data (including food assistance programs and other forms of nonprofits). However, putting the information to use is more difficult. Fewer than half (47%) say analyzing the data is easy, resulting in a slew of challenges when it comes to monitoring and quantifying things like effect and efficiency. Furthermore, only 41% believe it is simple to use data to improve the overall effect of programs. Despite the fact that charities are becoming more mobile with each passing year, only 29% claim they can easily capture and view data using a mobile device. This highlights the value of analytics tools for nonprofits, but according to the survey, only 45 percent of nonprofits actually use analytics, with another 30 percent aiming to do so over the next two years. Nonprofits have vast datasets that they can use to build mathematical models that will aid in fundraising optimization. In the current research pattern, this emphasis is absent. They will develop a more refined marketing plan by using appropriate visual data analytics to recognize and target the right groups. Data insights assist NGOs in identifying and categorizing donors based on a number of variables, helping them to better focus their marketing and fundraising efforts. Data analysis also assists charities in identifying partnerships that can be used to establish particular incentives. As a result, visualization tools are important for improving the presentation of narrative or verbal data and analysis in humanitarian relief operations.

Visual management is an organization system that strives to improve organizational performance by bring into line the vision, fundamental values, objectives and organizational culture with other management systems, work processes and elements, and finally stakeholders (Vilarinho et al., 2018). Dashboards can hence be defined as a graphical user interface that comprehends measures of various levels of performance to aid decision making. Yigitbasioglu and Velcu (2012) present a more complete definition: a visual and immersive performance assessment platform that shows the most relevant details for achieving one or more individuals and/or corporate goals on a single screen, developing the user to recognize, discover, and communicate problem areas that need corrective action. It is possible to encourage others to participate in the improvement process using this method. (Bititci et al., 2016) classify dashboards based on the level (strategic, tactical, or operational) and premise (planning or progress) for which they are designed. A transversal approach is developed in their research, which incorporates visual management at the strategic and operational levels. According to the authors, visual management methods at the strategic level are generally static, with an emphasis on contact from top to bottom. Visual approaches often carry some risk, such as the possibility of visual information being misinterpreted. Therefore, the design stage is considered a critical phase to tackle possible visualization risks and to make them effective. Several research methods have been conducted on the importance of panel design, development and execution. Pauwels et al. (2009) clarify what dashboards are, how should they be developed, the drivers and the obstacles to their adoption, recognizing the applicability of dashboards in all areas of an organization to support decision making. Their work emphasizes the importance of dashboards, given the rapid growth of dashboard implementation in large organizations. Allio (2012) describes common challenges in strategic dashboard design and implementation and offers recommendations for improving dashboard design, reach, usage, and effect. The author focuses on dashboards that help managers capture relevant data in order to optimize business plan execution. Yigitbasioglu and Velcu (2012) conducts a literature review on the use of dashboards as decision support tools in performance management and identifies potential design problems that organizations seeking to create and incorporate dashboards should consider. The authors argue that determining the goals of dashboards allows for practical adjustments to their functionality (visual and functional). Functional features allow a cognitive adjustment with different types of users, and visual characteristics allow to improve the process of visualization and interpretation of information. Finally, the features of the dashboard enable visual interpretation and information decoding which promotes decision support in performance management. The dashboard, according to Yigitbasioglu and Velcu (2012), should be interactive and versatile, taking into account the various purposes for which it is intended. They also warn that once dashboard users use presented data to help their decisions, they must be mindful of the impact of information overload. As a result, dashboard contents that divert user's attention away from the most important dashboard data should be avoided.

While information about the creation and implementation of dashboards can be found in the literature, it is generally focused on large organizations, especially for-profit organizations with advanced information systems. Hunger-relief organizations vary from large corporations in terms of contact and information processes, as well as other characteristics such as those described earlier, all of which have an impact on the dashboard creation process. Furthermore, these strategies are mostly aimed at executives and managers. Creating dashboards necessitates addressing problems that are directly related to organizational capacities or system maturity. Understanding the strengths and shortcomings of an organization's core capabilities involves assessing the degree of sophistication in terms of data handling and information systems. As a consequence, this information will assist in the detection of change and innovation actions, affecting the creation of dashboards. As a result, dashboards geared toward food banks that reflect on their unique characteristics, such as maturity, are missing in the literature. As a result, the proposed protocol and its implementation in food bank supply chain planning is a contribution to this sector.

Administrators of food banks face difficulties in making educated choices about food and resource allocations in order to fulfill the hunger needs of the people they represent. Despite the fact that food banks accept contributions from a variety of outlets, their ability to satisfy hunger is based on a number of factors. These can be due to supply (donations received) and demand (satisfying hunger) variations. This confusion may be caused by a lack of information about food donations, logistical issues, a lack of manpower, or insufficient decision-making resources. This research aims to combine data analytics, visualization and interactivity and provide a visual analytics framework to assist food bank administrators manage an effective food relief program.

6.3 Dashboard development process

The product development process proposed by Pahl and Beit (2013) has been recognized worldwide as one of the most systematic-based approaches to developing a product and it is this approach that we utilize to develop our proposed dashboard. In the product development process provided by Pahl and Beitz (2013), the authors have provided a detailed outline of the main stages in the product development process. This includes beginning the process with the understanding of the task that the product is developed for, followed by the conceptual design of the product, leading to the final prototype and the detailed design of the product. In the task understanding phase, steps are taken to collect information regarding the main requirements that the given product has to achieve keeping in mind the necessary constraints and the respective impacts. The conceptual design phase and the following phases after, are developed based on the main requirements which are continuously improved and updated. The phase of conceptual design basically controls the main concept of which the overall layout of the system boundary is obtained to get the final design phase. The final design phase is the final step of the product development consisting of the fabrication and handling procedures. Based on this approach, we have defined the required phases needed for product, which in our case is the dashboard, intended for hunger relief organizations such as food banks. The phases are explained below and also provided in Figure 6.1:

- *Gather supply and demand information and uncertainty sources for analysis* –
to gain a better understanding of the current state of the food bank under investigation, to determine priority improvement measures, and to gather feedback from employees and stakeholders for the dashboard;
- *Dashboard necessities evaluation* – to explain in a reasonable manner the requirements that the dashboard must meet, provided the data sources, literature, stakeholders, and project team expertise;
- *Dashboard layout development* – to translate defined needs and requirements into technical solutions, conducting and refining layouts until the most satisfactory outcomes are achieved;
- *Dashboard execution and improvement* – to bring the concept dashboard and the tools generated for its execution to the test, assess their performance, and make improvements.

Each step/phase should be meticulously prepared, carried out, and assessed. The outcomes of one step serve as the starting point for the next. As a result, the goals and key steps to take in each process are outlined below.

Figure 6.1 Defined phases for the dashboard development process

6.3.1 Gather supply and demand information and uncertainty sources for analysis

The existing state of the food bank should be assessed in order to identify potential opportunities for development. This step aims to collect dashboard recommendations and identify the key questions that should be taken into account. Preparation, implementation, outcome interpretation, and synthesis are the key stages of this step, and are close to the stages of the audit process. The gathering of knowledge about the organization to identify the organizational setting, priority demands, specific goals, the services and procedures that food banks have, as well as the scale and organizational structure are all part of the planning stage. This first stage also involves the establishment of a scope, the identification of specific divisions and parts to be included, and the selection of partners to join the project team. Stakeholders should be chosen based on their desire to add new insights, experiences, and knowledge to the dashboard development process, as well as their willingness to participate in dashboard implementation in the future. Finally, holding interviews to gather accurate knowledge about participants' perspectives and memories is part of the preparation stage. In order to learn about and comprehend the technologies, correspondence, and information processes in operation, the organization's system awareness should also be learned.

Following the data compilation and analysis, a synthesis should be performed, focusing on the flaws, growth opportunities, and dashboard recommendations.

6.3.2 Dashboard necessities evaluation

The aim of identifying dashboard specifications is to consider the needs of the stakeholders and ensure that they are satisfied with the final product. The dashboard should be created ahead of time to prevent discrepancies in stakeholders' expectations and the finished product. The findings from the first step, as well as the team's expertise and knowledge gathered from the literature, should all be taken into account when defining its characteristics. The obvious specifications are specified and reported first, followed by refinement. Refining requirements entails defining the goals that the solution must achieve as well as the properties that the dashboard must provide and can do without. A dashboard specification list should be organized clearly and take objections and changes into account (Pahl & Beitz, 2013). Intent, user characteristics, graphic characteristics, functional characteristics, contents, and decision-making should all be taken into account when planning specifications (Yigitbasioglu & Velcu, 2012). This artifacts should not be overly formalized; they are just a means of ensuring that key issues are not missed and offering supporting reference material. In order to produce good outcomes, the strategic objectives and targets, indicators and purposes, methods and services, and some of them as part of the dashboard contents should all be discussed (“Performance Dashboards: Measuring, Monitoring, and Managing Your Business - Wayne W. Eckerson - Google Books”, n.d.).

6.3.3 Dashboard layout development

The key aim of this process is to turn the specifications into technological solutions, which are provided through a dashboard interface and supporting resources, especially the data source that feeds the data into the dashboard. The definition and architecture are described in this process. Following the completion of the concept, many prototype layouts are created in order to gather more details about various options. The best interface design can be accomplished by refining and optimizing formats and reviewing technical requirements (Pahl & Beitz, 2013). Benchmarking, fundamental science, and innovation are some examples of methods that can help with dashboard interface growth. It's critical that the dashboard's features (visual and functional) and contents are appropriate for the dashboard's intent, as well as taking into account the user's knowledge and characteristics. Dashboard functionality must be specified in order to successfully facilitate visual perception and information interpretation.

Performance metrics that indicate the execution of an operation are typically presented on dashboards. Leading and slacking indicators are the two main categories of indicators. Leading metrics monitor actions that have a direct impact on potential performance. Many financial metrics, for example, are slacking indices that calculate the contribution of previous operations (Yigitbasioglu & Velcu, 2012). In order to ensure a mix of historical, present, and future success metrics, different types of indicators can be used in a dashboard ("Key Performance Indicators: Developing, Implementing, and Using Winning KPIs - David Parmenter - Google Books", n.d.). Metrics relevant to identifying supply and demand targets are recommended as a starting point for the set of steps that will be shown in the dashboard, and they can be viewed at various time periods (weekly, monthly, or yearly) to include visualization at various levels of preparation (strategic, tactical, or operational). Dashboards are often designed to be a long-term infrastructure that introduces, facilitates, and coordinates operations using freely accessible methodologies and tools. Tableau is a data visualization and research technology that has been widely applicable to food bank managers. This has been made available to the food banks free-of-cost thanks to the software grant received by Feeding America (Tableau, 2020).

6.3.4 Dashboard execution and improvement

Following the completion of the dashboard specification, the required supplies for dashboard construction and installation must be developed and implemented. This process also includes checks to ensure that the dashboard, its supporting software, underlying operations, and management activities are working as planned. This stage ensures that more functional versions are made. Measures should be made to ensure that correct and well organized documents are maintained in order to feed reliable data into the dashboard. This necessitates a thorough review of existing records in order to identify discrepancies and devise solutions. The documents should be correctly entered into the food bank's database system. Consider a data source that feeds data into the dashboard and into the information system. This enables proper data collection and management, as well as simpler synchronization and monitoring of data presented on the dashboard. The major dashboard management tasks can be clarified in a document that summarizes the key factors: accountable, upgrade times, activities to be done, necessary inputs, and desired outputs. In addition, at this point, any enhancement measures that will be enabled by the dashboard should be evaluated in order to identify and secure the tools and processes that will enable them to be executed effectively. Finally, the feature for generating information to be shown on the dashboard should be checked during this process. The relevance of the data, as well as the dashboard management tasks and the activities to be completed with the dashboard's assistance, should all be evaluated. This practice also includes defining and introducing enhancement methods.

Figures 6.2–6.5 depict each stage of the conceptual procedure, and are ordered according to the IDEF0 standard (Menzel & Mayer, 1998): input on the left, performance on the right, tools for completing the mission on the bottom, and targets on the top.

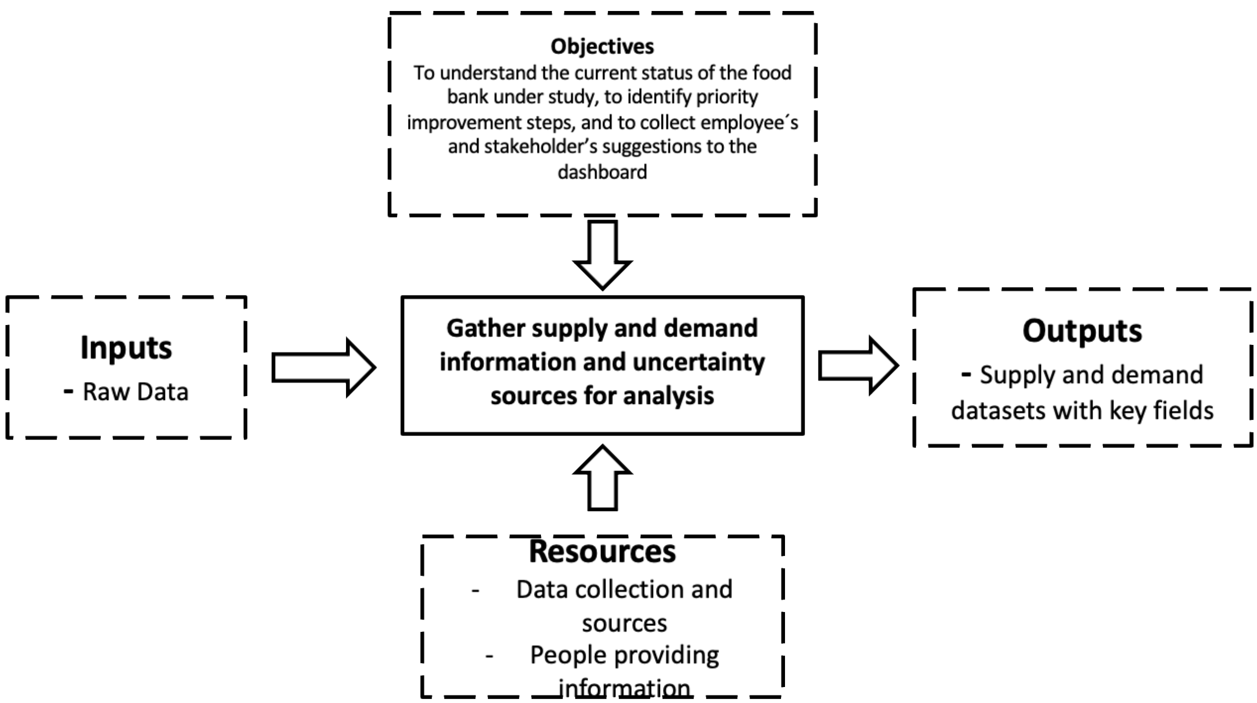


Figure 6.2 IDEF0 diagram of the first phase of the conceptual framework proposal

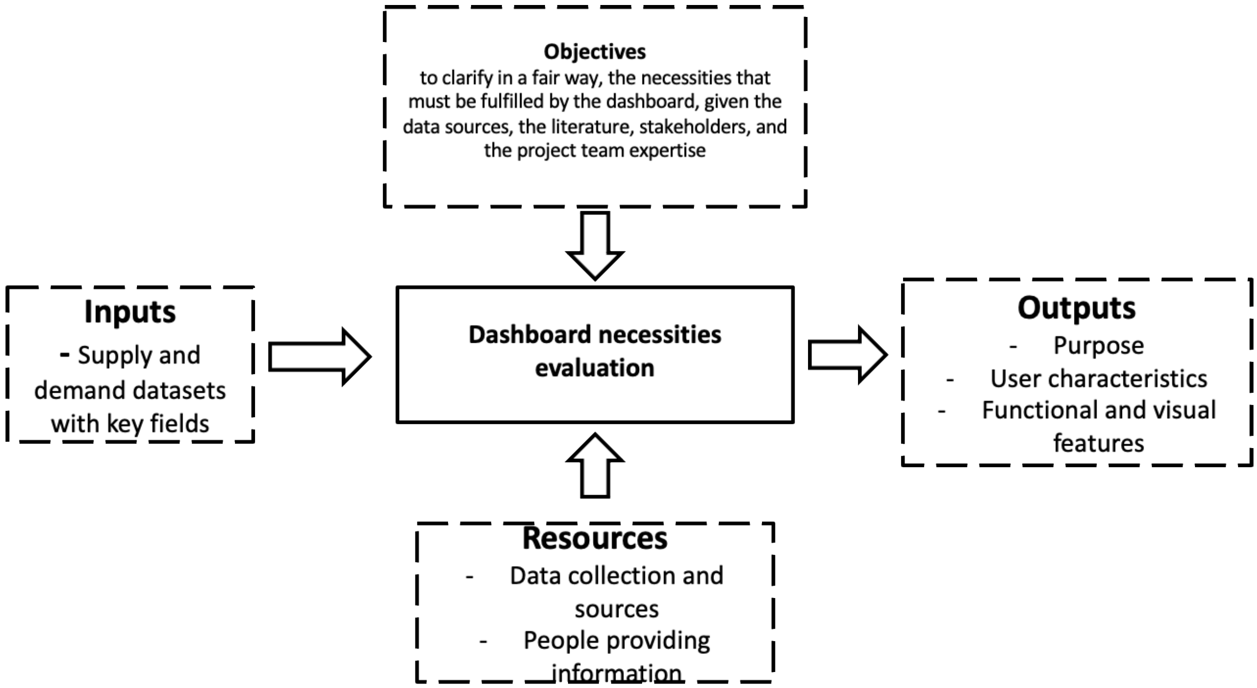


Figure 6.3 IDEF0 diagram of the second phase of the conceptual framework proposal

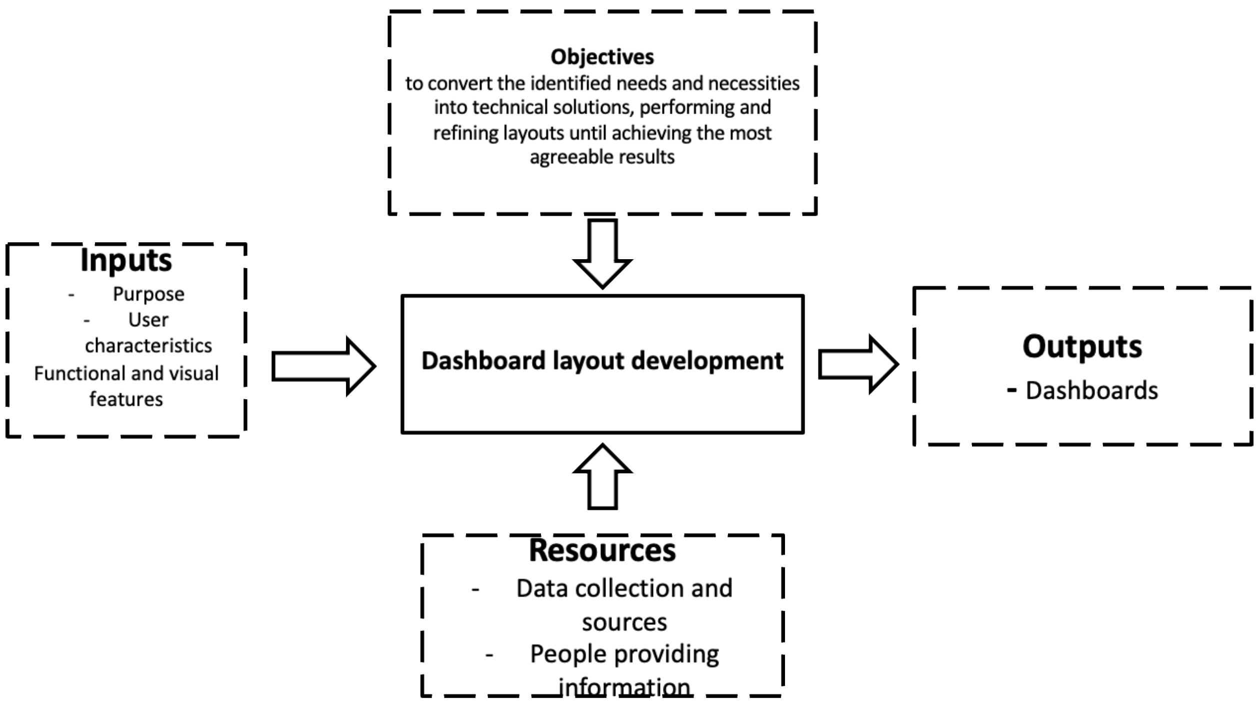


Figure 6.4 IDEF0 diagram of the third phase of the conceptual framework proposal

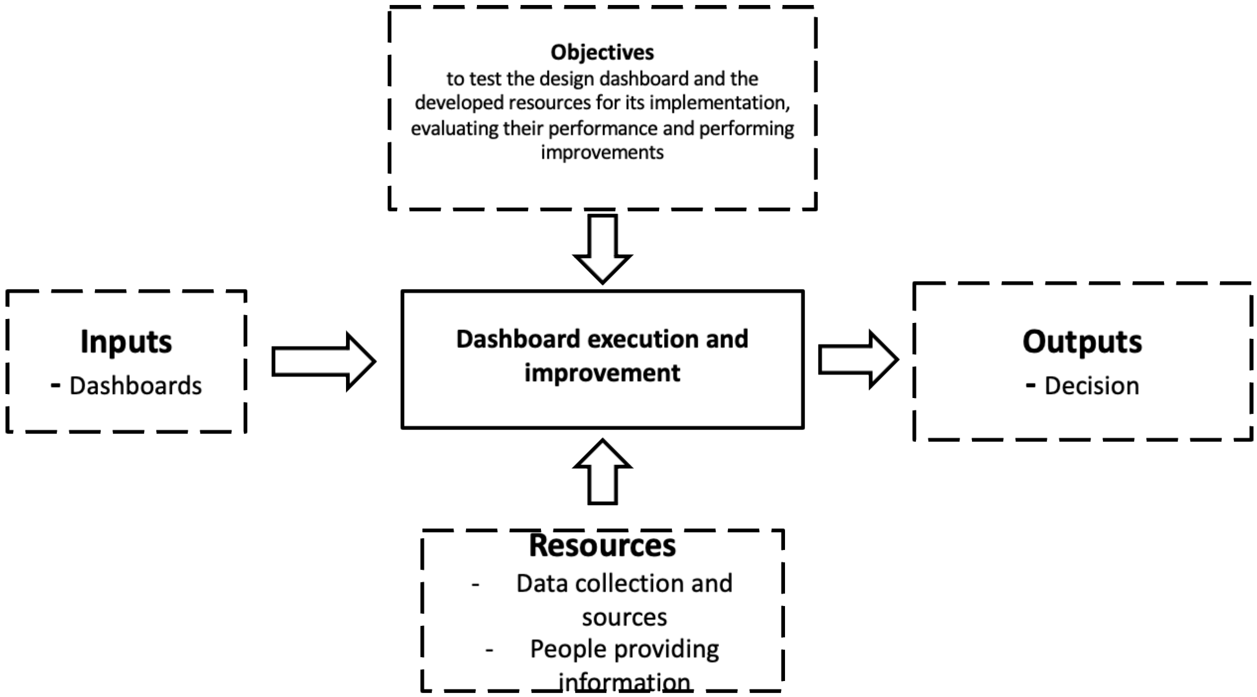


Figure 6.5 IDEF0 diagram of the fourth phase of the conceptual framework proposal

6.4 Framework implementation in a food bank organization

The pilot experiment took place at a community food bank that supports and manages over 180 food agencies, resulting in food being distributed to the hungry in 14 counties surrounding their agency. The company hires 14 employees, including two senior managers, two middle managers, eight warehouse workers, and two miscellaneous workers who perform one or two shifts a day. The following subsections, which conform to the established context, outline the execution of the proposed protocol on the regional food bank to assist food bank managers at different stages of preparation to manage a successful hunger-relief program.

6.4.1 Gather supply and demand information and uncertainty sources for analysis

The following topics were grouped in a diagnostic synthesis: origins of food donations, commodity forms of food donations, types of storage structures in food donation warehousing, types of food deprived population attending the food bank, location of food agencies, and dashboard recommendations. The data was organized into supply and demand information streams, and their performance metrics, inputs, outputs, services, information, events, and methods for performing process assessments, reporting, and analysis were examined. Food bank managers monitor food contributions on a weekly, monthly, and annual basis based on donor type (manufacturers, suppliers, restaurants, and others), food type (bread and cereals, fruits and vegetables, meats, dairy products, and confectionaries), and storage type (dry, refrigerated, and frozen). Administrators monitor demand based on the number of people who visit the food bank on a weekly, monthly, and annual basis for the demand visualization. Individuals are classified into three classes depending on the number of infants, elderly, and adults who attend the food bank and their respective agencies. It was decided that knowledge on supply and demand uncertainty at the strategic, tactical, and operational stages of preparation can be strengthened to enable clearer and more efficient planning that can be visualized by personnel at different levels of hierarchy, activity, and regulation of their supply chain and resources. At any point of planning, current performance metrics should be introduced and supplemented to be more compatible with the established processes' objectives. It was confirmed that information is not uniform across different levels of preparation in terms of resources and information. Based on the data sources presented, it was decided that a structured process for generating records that ensure the necessary variables for visualization is required. However, since these data sets are not created in a consistent and standardized manner, decisions may be made based on inaccurate and inconsistent evidence. As a result, steps should be taken to strengthen the information infrastructure so that accurate data can be collected from fruitful fields and converted into usable information for improved decision-making. Informal contact exists between executive leaders, middle management, and staff. On a weekly and monthly basis, middle managers and personnel meet to review the supply chain operation, and on a monthly and semi-annual basis, the management committee and the rest of the staff meet to discuss the strategic priorities. The need for efficient forecasting of food donation supply is also discussed, owing to the fluctuations of demand and supply of donated foods. As a result, the layout of the dashboard would take into account the forecasting of food donations based on the data segregations issued.

In addition to these results, recommendations for the dashboard were listed. The design and revision of cleaned data record sheets was proposed, enabling staff to record specific data for the management of the visualization's effectiveness. A review of the completed diagnosis was circulated to relevant individuals within the community and related to the initiative, with the intention of using it to help understand the present status of the food bank and its future path. According to the IDEF0 norm, Figure 6.2 depicts this process (Menzel and Mayer 1998).

6.4.2 Dashboard necessities evaluation

The criteria were calculated based on the findings of the diagnostic phase, the expertise of the team members, and the knowledge gathered from the literature. Table 6.1 outlines the key specifications based on the following criteria: intent, user characteristics, technical and visual capabilities, content, and decision-making support. In addition, Figure 6.3 depicts this step in accordance with the IDEF0 norm (Menzel & Mayer, 1998).

Table 6.1 Synthesis of identified requirements for the dashboard

Purpose	User characteristics	Functional and visual features	Contents	Support in decision-making
To have systematic data for decision-making at different stages of food delivery preparation	The information on the dashboard, as well as the planning hierarchy, should be simple to understand for all	To ensure comprehension and user engagement, the dashboard structure and visualization tools should take into account the sophistication of the information system and display the information in a concise and attractive manner	For the food bank organization, information pertaining to the strategic, tactical and operational level of planning is available	Decision-making that is focused on the visualization of correct and credible data
To have forecasted supply and demand across a variety of time horizons for accurate food delivery preparation	The executive board, middle management, warehouse employees, and operators are all expected to use the system	Information is presented through bar charts, pie charts and trend charts	Observation of the availability and demand of donated foods in relation to the management's requirements	Access to information for the target audience in a timely manner, allowing them to respond in sensitive situations

Table 6.1 continued

Create an integrated dashboard to help in the decision-making process by applying effective analysis strategies for food bank results		To ensure easy readability and ergonomically sound demonstration of knowledge for all, the proposed visualization tool should adhere to the concepts of data visualization	Dashboard data focused on the respective planning stages on a yearly, annual, and weekly basis	Taking steps based on facts and personal experience
To record the supply and demand relationship trends over time			Users may use this information to consider the actual and forecasted supply and demand scenario, as well as the impact, and to make effective planning decisions and use problem-solving tools.	

6.4.3 Dashboard layout development

The defined specifications were translated into technical solutions in this process, which were provided through dashboard layouts and drafts of the support tools that produce the information to be displayed on the dashboard. Initially designed layouts have been updated and refined to adapt to new concepts and changes. These ideas and changes were mainly due to the previous literature and the need of stakeholders, particularly senior managers, warehouse workers and middle managers who provide input and are decision-makers for their respective planning management systems. The changes made also took into account the tools and knowledge available at the food bank at the time of the dashboard's creation. Forecasting models from the literature were introduced into dashboards for each stage of preparation to help reduce the variability of food availability and demand.

As a result, each dashboard interface has two main zones (food supply study zone and food demand study zone), which are specified based on the frequency of information updates and the quality of the contents. In turn, each zone can be subdivided into several areas specified by the function of the contents. The three main dashboards are titled Strategic level of planning, Tactical level of planning, and Operational level of planning. The strategic stage of strategy dashboard (Fig 6.3) offers a yearly summary of the current supply and demand for donated foods. Food supply has been classified according to food type, storage type, and donor type. To view discrete data in the respective sample areas, pie charts and bar charts are given as visualization resources. The tactical stage of forecasting dashboard (Fig. 6.4) shows supply and demand on a monthly basis. For successful tactical preparation, the supply and demand outlook for the next month is also listed. The operational stage of the planning dashboard (Fig. 6.5) displays a weekly view of supply and demand, as well as estimates for the coming weeks. The aim of the Trends section of the food distribution is to see how things have changed over the months or weeks. The aim of the Current Status section is to aid in the more thorough review of such cases. Overall, we created dynamic dashboards that combine data analytics and visualization strategies to ensure accessibility and help in successful decision-making. According to the IDEF0 norm, Figure 6.4 depicts this process (Menzel & Mayer, 1998).

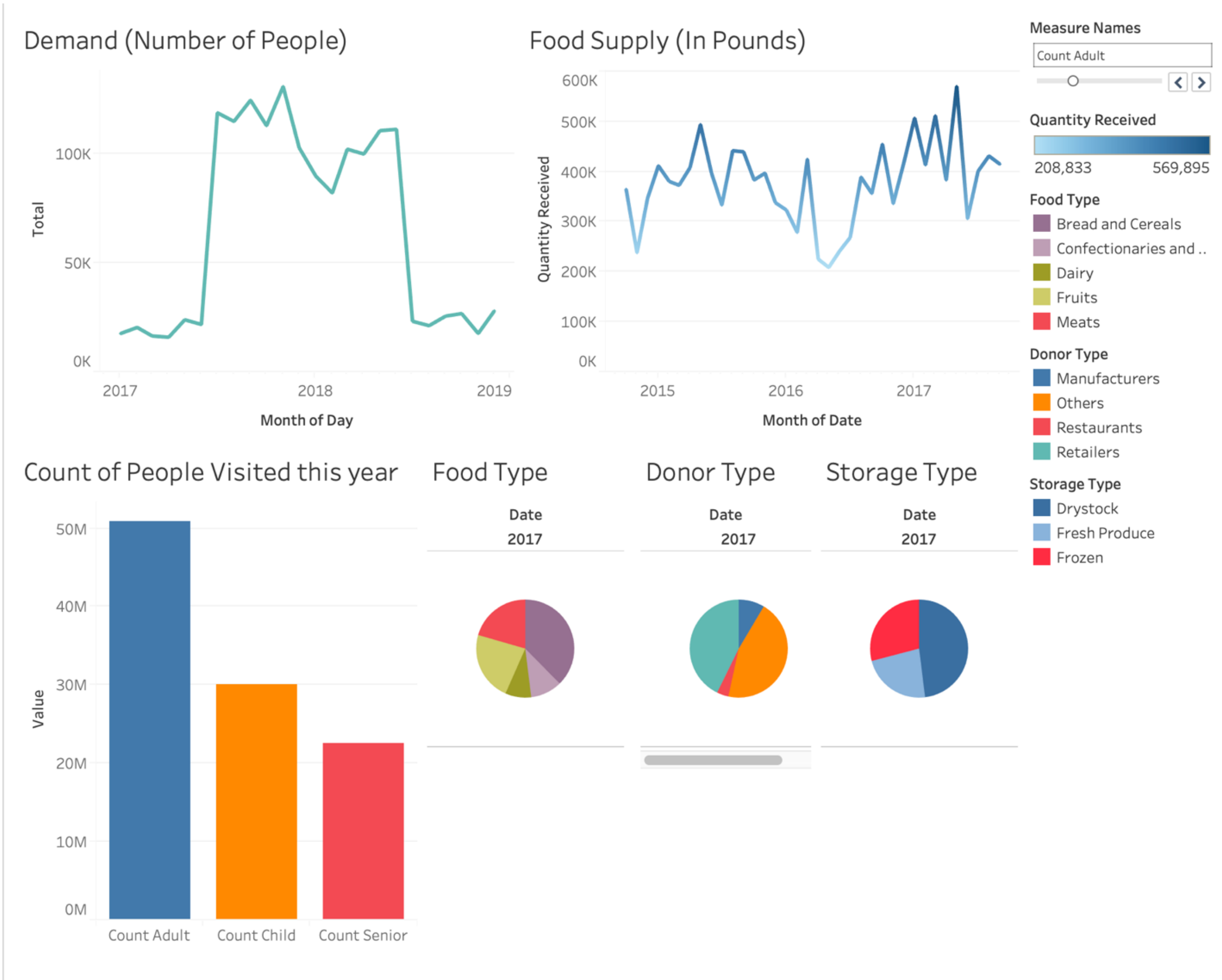


Figure 6.6 Strategic level of planning dashboard

Figure 6.7 Tactical level of planning dashboard

Figure 6.8 Operational level of planning dashboard

6.4.4 Dashboard execution and improvement

The data is presented in the form of an Excel ® spreadsheet deliberately designed to aggregate the data used to produce updated information, as well as the actual and forecasted effects of the given information for each stage of preparation and their respective time periods of forecasting. To ensure quality development and to assist the user in interpreting and maintaining the worksheets, the data file is used to create data visualizations in different Tableau worksheets that are structured depending on the respective planning level. Any collaborator with basic programming knowledge should be able to understand and use the Excel ® file. It also aims to be functional and fast at upgrading data, as well as converting the data into useful information. Finally, the content of the worksheets, especially the basic worksheets used in the dashboards, can be quickly tailored to the needs of the food bank and new ideas from stakeholders.

In a final step, the following information was defined for each zone of the dashboard: the staff in charge of the preparation stage, upgrade periods, appropriate inputs, and forecasted outputs. This practice serves as a reference for dashboard managers, allowing for the detection of flaws and opportunities for change during the whole process of using the dashboard. Tableau is a simple web-based visualization platform that encourages active sharing of analysis results without requiring specialized technical skills. According to the IDEF0 norm, Figure 6.5 depicts this process (Menzel & Mayer, 1998).

6.5 Conclusions

Current platform development processes are typically geared toward multinational corporations with sophisticated computer structures and are often too technical for non-profit hunger relief organizations. Food banks have a lower degree of production and information technology sophistication than major corporations. To create proper and successful dashboards, this and other performance characteristics should be considered during dashboard creation. A procedure for developing dashboards in the sense of hunger relief was discussed in this article. The protocol aids these non-profits in developing appropriate dashboards for successful decision-making. The proposed process is divided into several stages: production area diagnosis, dashboard requirements evaluation, dashboard interface creation, dashboard implementation, and dashboard enhancement. This methodical approach was used to revise and improve the dashboards that were developed. This turned out to be a key step in ensuring the dashboard's required inputs. Based on the defined dashboard criteria, it was clear that the company desired a dashboard that combined the three types: operational, tactical, and strategic. The dashboard design is focused on product development criteria, prior literature, and stakeholder requirements. At different stages of preparation, the digital dashboards built would provide food bank managers with quick access to data, better knowledge reporting, and personalized data visualization. A data file that aggregates data used to create and update the information to post on the dashboard was used as a support tool. Finally, core reasons for dashboard management were identified, resulting in a dashboard management guide. This procedure was efficient and beneficial in creating a dashboard that was tailored to the reality of the food bank's priorities. Since it reflects on the specificities of food banks, the procedure outlined in this paper to build and create dashboards differs significantly from other evaluation frameworks available in the literature. Furthermore, the protocol incorporates planning techniques and data visualization processes, resulting in significant advances in fair and efficient food delivery for those who are hungry. The technique may be extended to other food banks or international aid organizations with the same aim due to its flexibility and comprehensiveness. In the future, it is hoped to expand and implement the dashboard to other food banks, as well as consult with them to assess the dashboard's feasibility and make revisions based on their feedback, which may lead to a more comprehensive iteration of the proposed protocol. Future study may perform in-depth interviews or surveys focused on the same questionnaires to discover additional gaps in visualization techniques.

7. CONCLUSION AND FUTURE RESEARCH

In this dissertation, various machine learning and statistical approaches have been studies for understanding the dynamic behavior of the supply and demand of the food donations in regional food banks.

In Chapter 2, through a systematic and structured literature review, insights were provided into the conceptualization and modelling ideas on the issues related to fairness and sustainability, cost reduction, food quality and nutrition, and data uncertainty. In this perspective, total 48 previous published research articles were selected, categorized, and reviewed to find the research gaps and future scope of research. The research gaps were identified, discussed, and suggestions were made for future research opportunity.

In Chapter 3, a novel hybrid ARIMA-NNAR model with the implementation of Walk-forward cross validation was implemented to study the univariate analysis of the food supply dataset that explains the linear and nonlinear autocorrelation structures present in the data better than the traditional component.

In Chapter 4, generalized data-driven framework is proposed to assess the supply estimate and to identify the key predictors in the model. Five parametric, semi-parametric and non-parametric machine learning techniques were evaluated: GLM. GAM, SVM, RF, and BART to explore the patterns in the supply of donated food and purchased food process in the food bank organization. The results suggest that Random Forest (RF) outperforms all other tested statistical methods in terms of capturing the supply estimate for both donated and purchased goods

In Chapter 5, the results of the food accessibility pattern study, food demand of the a regional food bank organization is studied by developing predictive models. It is seen that by implementation of clustering results to the predictive models have an obvious accuracy improvement and hence a two-stage hybrid demand estimation model is proposed based on the results obtained.

In Chapter 6, a procedure to develop dashboards in the context of hunger-relief was presented. The procedure assists these non-profit organizations to achieve suitable dashboards for effective decision-making. The proposed procedure consists of several phases: diagnosis of the production area; dashboard requirements assessment; dashboard layout development; dashboard implementation and improvement. This systematic approach was implemented to revise and optimize the dashboards developed.

The future work to extend this research include:

- Extending the systematic literature review to analyze the work focused extended for the study of the food bank logistics on a systems perspective
- Behavior of the proposed hybrid time series model in Chapter 3 for seasonal and multivariate time series datasets can be considered as a future research work
- Involvement of these predictive models with optimization models such as routing and inventory management to ensure optimal allocation of food to the people in need
- develop and implement the dashboard to other food banks and collaborate with them to evaluate the effectiveness of the dashboards and revise it based on their subjective that could lead to a more robust version of the proposed procedure
- Future studies can undertake extensive interview or surveys based on the specific questionnaires, to identify other differences in the visualization tools

8. REFERENCES

Aboujaoude, M., & Schneider, K. (2017). A Decision Support Tool for Scheduling Food Bank Compliance Audits. *IIE Annual Conference. Proceedings,* 356-361. https://search.proquest.com/scholarly-journals/decision-support-tool-scheduling-food-bank/docview/1951123944/se-2?accountid=13360

Adivar, B., Atan, T., Oflaç, B. S., & Örten, T. (2010). Improving social welfare chain using optimal planning model. *Supply Chain Management: An International Journal*. https://doi/10.1108/13598541011054661

Agresti, A. (2003). *Categorical data analysis* (Vol. 482). John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=hpEzw4T0sPUC&oi=fnd&pg=PR7&dq=Categorical+Data+Analysis+-+Alan+Agresti+-+Google+Books&ots=nsBlalhtTa&sig=sL0gKskS6MrBvl3cCVX09Uuxxyk#v=onepage&q=Categorical%20Data%20Analysis%20-%20Alan%20Agresti%20-%20Google%20Books&f=false

Ahire, S. L., & Pekgün, P. (2018). Harvest hope food bank optimizes its promotional strategy to raise donations using integer programming. *Interfaces*, *48*(4), 291-306. https://doi.org/10.1287/inte.2018.0944

Aiello, G., Enea, M., & Muriana, C. (2015). Alternatives to the traditional waste management: food recovery for human non-profit organizations. *International Journal of Operations and Quantitative Management*, *21*(3), 215-239. https://www.researchgate.net/publication/301589046

Allio, M. K. (2012). Strategic dashboards: designing and deploying them to improve implementation. *Strategy & Leadership*. https://doi.org/10.1108/10878571211257159

Alotaik, O., Chellappa, V. S., Steadman, S., & Kumara, S. (2017). Little data, big data, and food insecurity: An integrated approach. In *IIE Annual Conference. Proceedings* (pp. 1490-1495). Institute of Industrial and Systems Engineers (IISE). https://search.proquest.com/docview/1951123178/fulltextPDF/B9171A34E8954033PQ/1?accountid=13360

America, F. (2014). Hunger in America 2014. National Report. August. https://phcnpg.org/docs/Welcome%20Page/hunger-in-america-2014-full-report.pdf

Amorim, M., Ferreira, S., & Couto, A. (2018). Emergency medical service response: analyzing vehicle dispatching rules. *Transportation research record*, *2672*(32), 10-21. https://doi.org/10.1177/0361198118781645

An, R., Wang, J., Liu, J., Shen, J., Loehmer, E., & McCaffrey, J. (2019). A systematic review of food pantry-based interventions in the USA. *Public health nutrition*, *22*(9), 1704-1716. https://doi.org/10.1017/S1368980019000144

Announcing the Latest Nonprofit Trends Report. (2020, February 18) Salesforce.org. Retrieved August 28, 2020, from https://www.salesforce.org/blog/nonprofit-trends-report-second-edition-research/

Arora, S., & Taylor, J. W. (2016). Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, *59*, 47-59. https://doi.org/10.1016/j.omega.2014.08.008

Ata, B., Lee, D., & Sönmez, E. (2019). Dynamic volunteer staffing in multicrop gleaning operations. *Operations Research*, *67*(2), 295-314. https://doi.org/10.1287/opre.2018.1792

Bacon, C. M., & Baker, G. A. (2017). The rise of food banks and the challenge of matching food assistance with potential need: towards a spatially specific, rapid assessment approach. *Agriculture and Human Values*, *34*(4), 899-919. https://doi.org/10.1007/s10460-017-9783-y

Balcik, B., Beamon, B. M., & Smilowitz, K. (2008). Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems*, *12*(2), 51-63. https://doi-org.ezproxy.lib.purdue.edu/10.1080/15472450802023329

Balcik, B., Iravani, S., & Smilowitz, K. (2014). Multi-vehicle sequential resource allocation for a nonprofit distribution system. *IIE Transactions*, *46*(12), 1279-1297. https://doi.org/10.1080/0740817X.2013.876240

Balza-Franco, V., Paternina-Arboleda, C. D., Cantillo, V., Macea, L. F., & Ramírez-Ríos, D. G. (2017). A collaborative supply chain model for non-for-profit networks based on cooperative game theory. *International Journal of Logistics Systems and Management*, *26*(4), 475-496. https://doi.org/10.1504/IJLSM.2017.082614

Battling Food Insecurity during a Pandemic. (2020, June 14) Providence. Retrieved August 28, 2020, from https://www.providence.org/news/uf/617667977

Bazerghi, C., McKay, F. H., & Dunn, M. (2016). The role of food banks in addressing food insecurity: a systematic review. *Journal of community health*, *41*(4), 732-740. https://doi.org/10.1007/s10900-015-0147-5

Beamon, B. M., & Balcik, B. (2008). Performance measurement in humanitarian relief chains. *International Journal of Public Sector Management*. https://doi.org/10.1108/09513550810846087

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70-83. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.csda.2017.11.003

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer. https://doi.org/10.1117/1.2819119

Bititci, U., Cocca, P., & Ates, A. (2016). Impact of visual performance management systems on the performance management practices of organisations. *International Journal of Production Research*, *54*(6), 1571-1593. https://doi.org/10.1080/00207543.2015.1005770

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brock III, L. G., & Davis, L. B. (2015). Estimating available supermarket commodities for food bank collection in the absence of information. *Expert Systems with Applications*, *42*(7), 3450-3461. https://doi.org/10.1016/j.eswa.2014.11.068

Bucknum, M., & Bentzel, D. (2019). Food Banks as Local Food Champions: How Hunger Relief Agencies Invest in Local and Regional Food Systems. In *Institutions as Conscious Food Consumers* (pp. 285-305). Academic Press. https://doi-org.ezproxy.lib.purdue.edu/10.1016/B978-0-12-813617-1.00013-7

Buisman, M. E., Haijema, R., Akkerman, R., & Bloemhof, J. M. (2019). Donation management for menu planning at soup kitchens. *European Journal of Operational Research*, *272*(1), 324-338. https://doi.org/10.1016/j.ejor.2018.06.005

Campbell, C. C. (1991). Food insecurity: a nutritional outcome or a predictor variable? *The Journal of nutrition*, *121*(3), 408-415. https://doi.org/10.1093/jn/121.3.408

Casey, P. H., Szeto, K., Lensing, S., Bogle, M., & Weber, J. (2001). Children in food-insufficient, low-income families: prevalence, health, and nutrition status. *Archives of pediatrics & adolescent medicine*, *155*(4), 508-514. https://doi.org/10.1001/archpedi.155.4.508

Çelik, M., Ergun, Ö., Johnson, B., Keskinocak, P., Lorca, Á., Pekgün, P., & Swann, J. (2012). Humanitarian logistics. In *New directions in informatics, optimization, logistics, and production* (pp. 18-49). INFORMS. https://doi.org/10.1287/educ.1120.0100

Chakraborty, T., Chattopadhyay, S., & Ghosh, I. (2019). Forecasting dengue epidemics using a hybrid methodology. *Physica A: Statistical Mechanics and its Applications*, *527*, 121266. https://doi.org/10.1016/j.physa.2019.121266

Chapnick, M., Barnidge, E., Sawicki, M., & Elliott, M. (2019). Healthy options in food pantries—a qualitative analysis of factors affecting the provision of healthy food items in St. Louis, Missouri. *Journal of Hunger & Environmental Nutrition*, *14*(1-2), 262-280. https://doi-org.ezproxy.lib.purdue.edu/10.1080/19320248.2017.1284027

Chellappa, V. S. (2016). *Spatial Accessibility in Food Deserts*. [Masters dissertation, Pennsylvania State University]. PennState repository. https://etda.libraries.psu.edu/files/final_submissions/13580

Lee, D., Sönmez, E., Gómez, M. I., & Fan, X. (2017). Combining two wrongs to make two rights: Mitigating food insecurity and food waste through gleaning operations. *Food Policy*, *68*, 40-52. https://doi.org/10.1016/j.foodpol.2016.12.004

Cordeiro, G. M., & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 629-643. https://www.jstor.org/stable/pdf/2345592.pdf?refreqid=excelsior%3Ab3e1a9e9427f99795fe483623cf19ec6

Davis, L. B., Jiang, S. X., Morgan, S. D., Nuamah, I. A., & Terry, J. R. (2016). Analysis and prediction of food donation behavior for a domestic hunger relief organization. *International Journal of Production Economics*, *182*, 26-37. https://doi.org/10.1016/j.ijpe.2016.07.020

Davis, L. B., Sengul, I., Ivy, J. S., Brock III, L. G., & Miles, L. (2014). Scheduling food bank collections and deliveries to ensure food safety and improve access. *Socio-Economic Planning Sciences*, *48*(3), 175-188. https://doi.org/10.1016/j.seps.2014.04.001

Delpish, R., Jiang, S., Davis, L., & Odubela, K. (2018, July). A visual analytics approach to combat confirmation bias for a local food bank. In *International Conference on Applied Human Factors and Ergonomics* (pp. 13-23). Springer, Cham. https://doi.org/10.1007/978-3-319-94391-6_2

Denyer, D., Tranfield, D., & Van Aken, J. E. (2008). Developing design propositions through research synthesis. *Organization studies*, *29*(3), 393-413. https://doi.org/10.1177/0170840607088020

Desai, Y. S. (2015). *Using visual analytics to support humanitarian relief* (Order No. 10003456) [Doctoral dissertation, North Carolina Agricultural and Technical State University]. ProQuest Dissertations & Theses Global. (1761843924). https://search.proquest.com/dissertations-theses/using-visual-analytics-support-humanitarian/docview/1761843924/se-2?accountid=13360

Divina, F., Garcia Torres, M., Goméz Vela, F. A., & Vazquez Noguera, J. L. (2019). A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings. *Energies*, *12*(10), 1934. https://doi.org/10.3390/en12101934

Domingos, S. D. O., de Oliveira, J. F., & de Mattos Neto, P. S. (2019). An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowledge-Based Systems*, *175*, 72-86. https://doi.org/10.1016/j.knosys.2019.03.011

Drackley, A. (2010). *How generous are we? Forecasting and demographic correlates of blood donation* (Identifier No. 458155992048567) [Doctoral dissertation, McMaster University]. MacSphere’s Institutional Repository. http://hdl.handle.net/11375/9460

Drackley, A., Newbold, K. B., Paez, A., & Heddle, N. (2012). Forecasting Ontario's blood supply and demand. *Transfusion*, *52*(2), 366-374. https://doi.org/10.1111/j.1537-2995.2011.03280.x

Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=5nuYDwAAQBAJ&oi=fnd&pg=PR7&dq=W.W.+Eckerson,+Performance+Dashboards:+Measuring,+Monitoring,+and+Managing+Your+Business,+Second+Ed&ots=AJIbwROrkx&sig=yovkiohS0AOc6d27j_hsdU0NxbY#v=onepage&q&f=false

Eisenhandler, O., & Tzur, M. (2019a). A segment-based formulation and a matheuristic for the humanitarian pickup and distribution problem. *Transportation Science*, *53*(5), 1389-1408. https://doi.org/10.1287/trsc.2019.0916

———. (2019b). The Humanitarian Pickup and Distribution Problem. *Operations Research 67 (1)*, 10–32. https://doi.org/10.1287/opre.2018.1751

Food Waste Loss and Donation. (2013, May). U.S. Environmental Protection Agency. Retrieved August 28, 2020, from https://www.epa.gov/sites/production/files/2016-01/documents/msw_task11_foodwastelossanddonation_508_fnl2.pdf

Feeding America. (n.d.). *What is a food bank?*. https://www.feedingamerica.org/our-work/food-bank-network#:~:text=A%20food%20bank%20is%20a,to%20people%20struggling%20with%20hunger

Feeding America's Foods to Encourage Background. (2015, July). Feeding America. Retrieved August 28, 2020, from https://hungerandhealth.feedingamerica.org/resource/foods-to-encourage-background/?_ga=2.203566096.1641490003.1505783484-2089554786.1505783484

Ferreira, G. O., Arruda, E. F., & Marujo, L. G. (2018). Inventory management of perishable items in long-term humanitarian operations using Markov Decision Processes. *International Journal of Disaster Risk Reduction*, *31*, 460-469. https://doi.org/10.1016/j.ijdrr.2018.05.010

Fianu, Sefakor, and Lauren B. Davis. "A Markov decision process model for equitable distribution of supplies under uncertainty." *European Journal of Operational Research* 264.3 (2018): 1101-1115. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.ejor.2017.07.017

Firat, M., Yurdusev, M. A., & Turan, M. E. (2009). Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water resources management*, *23*(4), 617-632. https://doi-org.ezproxy.lib.purdue.edu/10.1007/s11269-008-9291-3

Fraley, C., & Raftery, A. E. (2007). Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of classification*, *24*(2), 155-181. https://doi.org/10.1007/s00357-007-0004-5

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67. https://www.jstor.org/stable/pdf/2241837.pdf?refreqid=excelsior%3Abad6ee13354af2a82e11b2e7515e685e

Galli, F., Cavicchi, A., & Brunori, G. (2019). Food waste reduction and food poverty alleviation: a system dynamics conceptual model. *Agriculture and human values*, *36*(2), 289-300. https://doi.org/10.1007/s10460-019-09919-0

Gharehyakheh, A., & Sadeghiamirshahidi, N. (2018). A sustainable approach in food bank logistics. In *Proceedings of the International Annual Conference of the American Society for Engineering Management.* (pp. 1-8). American Society for Engineering Management (ASEM).
https://search.proquest.com/conference-papers-proceedings/sustainable-approach-food-bank-logistics/docview/2193095033/se-2?accountid=13360

Gharehyakheh, A., Krejci, C., Cantu, J., & Rogers, J. (2019, May). Dynamic Shelf-Life Prediction System to Improve Sustainability in Food Banks. In *Proceedings of the 2019 IISE Annual Conference, Orlando, FL, USA* (pp. 18-21). https://www.researchgate.net/publication/336749972

Glover, W. J., Poopunsri, T., & Hurley, R. (2014). Applying lean to non-profit organizations: A food bank case study. In *IIE Annual Conference. Proceedings* (p. 274). Institute of Industrial and Systems Engineers (IISE). https://search.proquest.com/docview/1622299475/fulltextPDF/6F1588D28EDE45A8PQ/1?accountid=13360

González-Torre, P. L., & Coque, J. (2016). From food waste to donations: the case of marketplaces in northern Spain. *Sustainability*, *8*(6), 575. https://doi.org/10.3390/su8060575

Gorr, W. L. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*, *10*(1), 1-4. https://doi.org/10.1016/0169-2070(94)90044-2

Gutjahr, W. J., & Fischer, S. (2018). Equity and deprivation costs in humanitarian logistics. *European Journal of Operational Research*, *270*(1), 185-197. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.ejor.2018.03.019

Hall, K. D., Guo, J., Dore, M., & Chow, C. C. (2009). The progressive increase of food waste in America and its environmental impact. *PloS one*, *4*(11), e7940. https://doi.org/10.1371/journal.pone.0007940

Hampl, J., & Hall, R. (2002). Dietetic approaches to US hunger and food insecurity. *Journal of the Academy of Nutrition and Dietetics*, *102*(7), 919-923. https://doi.org/10.1016/S0002-8223(02)90209-9

Handforth, B., Hennink, M., & Schwartz, M. B. (2013). A qualitative study of nutrition-based initiatives at selected food banks in the feeding America network. *Journal of the Academy of Nutrition and Dietetics*, *113*(3), 411-415. https://doi.org/10.1016/j.jand.2012.11.001

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC press. https://books.google.com/books?hl=en&lr=&id=qa29r1Ze1coC&oi=fnd&pg=PR13&dq=Hastie,+Trevor+J,+and+Robert+J+Tibshirani.+1990.+Generalized+Additive+Models.+BOOK.+Vol.+43.+CRC+Press.&ots=j5YSmqzVhQ&sig=45Mg2P5YMp9YaWD8YePCG9w25LM#v=onepage&q&f=false

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media. https://books.google.com/books?hl=en&lr=&id=tVIjmNS3Ob8C&oi=fnd&pg=PR13&dq=Hastie,+Trevor+J,+Robert+J+Tibshirani,+and+Jerome+H+Friedman.+2003.+The+Elements+of+Statistical+Learning:+Data+Mining,+Inference+and+Prediction.+Edited+by+Springer.+2nd+ed.&ots=EOCbQ7JZZ2&sig=YqMrt1usirYeAurzhB2PROgZVPU#v=onepage&q=Hastie%2C%20Trevor%20J%2C%20Robert%20J%20Tibshirani%2C%20and%20Jerome%20H%20Friedman.%202003.%20The%20Elements%20of%20Statistical%20Learning%3A%20Data%20Mining%2C%20Inference%20and%20Prediction.%20Edited%20by%20Springer.%202nd%20ed.&f=false

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, *268*(3), 836-851. https://doi.org/10.1016/j.ejor.2017.06.031.

Holguín-Veras, J., Jaller, M., Van Wassenhove, L. N., Pérez, N., & Wachtendorf, T. (2012). On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, *30*(7-8), 494-506. https://doi.org/10.1016/J.JOM.2012.08.003

How to Support Food Banks During Coronavirus. (n.d.). Feeding America. Retrieved July 20, 2020, from https://www.feedingamerica.org/take-action/coronavirus

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts. “Forecasting: Principles and Practice - Rob J Hyndman, George Athanasopoulos - Google Books.” https://books.google.com/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=R.J.+Hyndman,+G.+Athanasopoulos,+Forecasting:+Principles+and+Practice,+OTexts,+2018&ots=TieYyfVQNP&sig=nWhzt_JMifBqiijDdzxZ6YuT71k#v=onepage&q=R.J. Hyndman%2C G. Athanasopoulos%2C Forecasting%3A Principles and Practice%2C OTexts%2C 2018&f=false

Ibarra-Rojas, O. J., & Silva-Soto, Y. (2021). Vehicle routing problem considering equity of demand satisfaction. *Optimization Letters*, 1-23. https://doi-org.ezproxy.lib.purdue.edu/10.1007/s11590-021-01704-5

Irani, Z., Sharif, A. M., Lee, H., Aktas, E., Topaloğlu, Z., van't Wout, T., & Huda, S. (2018). Managing food security through food waste and loss: Small data to big data. *Computers & Operations Research*, *98*, 367-383. https://doi.org/10.1016/j.cor.2017.10.007

Jain, A., & Ormsbee, L. E. (2002). Short-term water demand forecast modeling techniques—CONVENTIONAL METHODS VERSUS AI. *Journal-American Water Works Association*, *94*(7), 64-72. https://doi.org/10.1002/j.1551-8833.2002.tb09507.x

Jiang, S., Davis, L., De Mleo, H. T., & Terry, J. (2013). Using data mining to analyze donation data for a local food bank. In *Proceedings of the International Conference on Data Science (ICDATA)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). http://worldcomp-proceedings.com/proc/p2013/DMI8075.pdf

Kaiser, M. L. (2011). Food security: an ecological–social analysis to promote social development. *Journal of Community Practice*, *19*(1), 62-79. https://doi.org/10.1080/10705422.2011.550261

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, *11*(2), 2664-2675. https://doi.org/10.1016/j.asoc.2010.10.015

Ko, I., & Chang, H. (2018). Interactive data visualization based on conventional statistical findings for antihypertensive prescriptions using National Health Insurance claims data. *International journal of medical informatics*, *116*, 1-8. https://doi.org/10.1016/j.ijmedinf.2018.05.003.

Kordanuli, B., Barjaktarović, L., Jeremić, L., & Alizamir, M. (2017). Appraisal of artificial neural network for forecasting of economic parameters. *Physica A: Statistical Mechanics and its Applications*, *465*, 515-519. https://doi.org/10.1016/j.physa.2016.08.062

Kumar, S., Niedan-Olsen, K., & Peterson, L. (2009). Educating the supply chain logistics for humanitarian efforts in Africa: a case study. *International Journal of Productivity and Performance Management*. https://doi.org/10.1108/17410400910965733

Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=MZgtBAAAQBAJ&oi=fnd&pg=PR15&ots=2D_XPbjU3U&sig=Ndo7ij94OPR6UYEsUsNEIiS44kg#v=onepage&q&f=false

Lee, D., Sönmez, E., Gómez, M. I., & Fan, X. (2017). Combining two wrongs to make two rights: Mitigating food insecurity and food waste through gleaning operations. *Food Policy*, *68*, 40-52. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.foodpol.2016.12.004

Lee, J. S., & Frongillo Jr, E. A. (2001). Nutritional and health consequences are associated with food insecurity among US elderly persons. *The Journal of nutrition*, *131*(5), 1503-1509. https://doi.org/10.1093/jn/131.5.1503

Lien, R. W., Iravani, S. M., & Smilowitz, K. R. (2014). Sequential resource allocation for nonprofit operations. *Operations Research*, *62*(2), 301-317. https://doi.org/10.1287/opre.2013.1244

Ma, Z., Li, H., Sun, Q., Wang, C., Yan, A., & Starfelt, F. (2014). Statistical analysis of energy consumption patterns on the heat demand of buildings in district heating systems. *Energy and Buildings*, *85*, 464-472. https://doi.org/10.1016/j.enbuild.2014.09.048

Maleki, A., Nasseri, S., Aminabad, M. S., & Hadi, M. (2018). Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics. *KSCE Journal of Civil Engineering*, *22*(9), 3233-3245. https://doi.org/10.1007/s12205-018-1195-z

Malone-Smolla, C. (2020, November 24). *Tableau powers Feeding America's digital advocacy for hunger relief*. Tableau. https://www.tableau.com/blog/2020/11/tableau-powers-feeding-americas-digital-advocacy-hunger-relief

Martins, C. J. (2019). *Redesign of a sustainable food bank supply chain.* [Doctoral dissertation, Universidade de Lisboa]. Repositório da Universidade de Lisboa. http://hdl.handle.net/10400.5/18139

Martins, C. L., Melo, M. T., & Pato, M. V. (2019). Redesigning a food bank supply chain network in a triple bottom line context. *International Journal of Production Economics*, *214*, 234-247. https://doi.org/10.1016/j.ijpe.2018.11.011

Mattogno, C., Ricci, M., Monardo, B., Valentino, P. A., & Palazzo, A. L. (2014). Feeding the City. Foodsheds and Urban Agriculture in San Diego. *CSE-City Safety Energy*, (2), 29-36. https://doi.org/10.12896/CSE20140020025

McCulloch, C. E. (2000). Generalized linear models. *Journal of the American Statistical Association*, *95*(452), 1320-1324. https://doi.org/10.1080/01621459.2000.10474340

McKay, F. H., Haines, B. C., & Dunn, M. (2019). Measuring and understanding food insecurity in Australia: A systematic review. *International journal of environmental research and public health*, *16*(3), 476. https://doi.org/10.3390/ijerph16030476

Mehtab, S., & Sen, J. (2020). A time series analysis-based stock price prediction using machine learning and deep learning models. *arXiv preprint arXiv:2004.11697*

Meline, T. (2006). Selecting studies for systemic review: Inclusion and exclusion criteria. *Contemporary issues in communication science and disorders*, *33*(Spring), 21-27. https://pubs.asha.org/doi/pdf/10.1044/cicsd_33_S_21

Menzel, C., & Mayer, R. J. (1998). The IDEF family of languages. In *Handbook on architectures of information systems* (pp. 209-241). Springer, Berlin, Heidelberg. https://doi-org.ezproxy.lib.purdue.edu/10.1007/978-3-662-03526-9_10

Merwe, S. V. D. (2018). Time Series Analysis of the Southern Oscillation Index using Bayesian Additive Regression Trees. *arXiv preprint arXiv:1804.01809*

Middleton, D. (1988). Mathematical statistics and data analysis, by John A. Rice. Pp 595.1988. ISBN 0-534-08247-5 (Wadsworth & Brooks/Cole). *The Mathematical Gazette*, *72*(462), 330-331. https://books.google.com/books?hl=en&lr=&id=EKA-yeX2GVgC&oi=fnd&pg=PR4&dq=rice+mathematical+statistics+and+data+analysis&ots=YhvtzThwH3&sig=CblvpBIFDv0LxctQX_cBd_RDDFw#v=onepage&q=rice mathematical statistics and data analysis&f=false

Mohan, S., Gopalakrishnan, M., & Mizzi, P. J. (2013). Improving the efficiency of a non-profit supply chain for the food insecure. *International Journal of Production Economics*, *143*(2), 248-255. https://doi.org/10.1016/j.ijpe.2011.05.019

Mousa, T. Y., & Freeland-Graves, J. H. (2017). Organizations of food redistribution and rescue. *Public health*, *152*, 117-122. https://doi.org/10.1016/j.puhe.2017.07.031

MyPlate & What Is a Serving of Fruits and Vegetables? Fill Half Your Plate. (n.d.). Fruitsandveggies. Retrieved December 13, 2019, from https://fruitsandveggies.org/stories/how-much-does-your-family-need/?gclid=CjwKCAiA58fvBRAzEiwAQW-hza9DwxeIVfvrVhOOOGik6kZmoxlmr1O1Ue5pVpgkovUjdZdB0oH84xoCmhEQAvD_BwE

Nair, D. J., Grzybowska, H., Rey, D., & Dixit, V. V. (2016a). Food rescue and delivery: Heuristic algorithm for periodic unpaired pickup and delivery vehicle routing problem. *Transportation Research Record*, *2548*(1), 81-89. https://doi.org/10.3141/2548-10

Nair, D. J., Rey, D., Dixit, V., & Valenta, T. (2016b). *Models for food rescue and delivery: Routing and resource allocation problem* (No. 16-2584). https://trid.trb.org/view.aspx?id=1392957

Nair, D. J., Rashidi, T. H., & Dixit, V. V. (2017a). Estimating surplus food supply for food rescue and delivery operations. *Socio-Economic Planning Sciences*, *57*, 73-83. https://doi.org/10.1016/J.SEPS.2016.09.004

Nair, D. J., Rey, D., & Dixit, V. V. (2017b). Fair allocation and cost-effective routing models for food rescue and redistribution. *IISE Transactions*, *49*(12), 1172-1188. https://doi-org.ezproxy.lib.purdue.edu/10.1080/24725854.2017.1351043

Nanney, M. S., Grannon, K. Y., Cureton, C., Hoolihan, C., Janowiec, M., Wang, Q., ... & King, R. P. (2016). Application of the Healthy Eating Index-2010 to the hunger relief system. *Public health nutrition*, *19*(16), 2906-2914. https://doi:10.1017/S136898001600118X

Nasr, G. E., Badr, E. A., & Younes, M. R. (2002). Neural networks in forecasting electrical energy consumption: univariate and multivariate approaches. *International Journal of Energy Research*, *26*(1), 67-78. https://doi.org/10.1002/er.766

Nelson, K., Cunningham, W., Andersen, R., Harrison, G., & Gelberg, L. (2001). Is food insufficiency associated with health status and health care utilization among adults with diabetes?. *Journal of General Internal Medicine*, *16*(6), 404-411. https://doi.org/10.1046/j.1525-1497.2001.016006404.x

Nuamah, I. A. (2016). *Predicting and Optimizing the Fair Allocation of Donations* (Order No. 10167912) [Doctoral dissertation, North Carolina Agricultural and Technical State University]. ProQuest Dissertations & Theses Global. (1845073012). https://search.proquest.com/dissertations-theses/predicting-optimizing-fair-allocation-donations/docview/1845073012/se-2?accountid=13360

Oliveira, M., & Torgo, L. (2015, February). Ensembles for time series forecasting. In *Asian Conference on Machine Learning* (pp. 360-370). PMLR. http://proceedings.mlr.press/v39/oliveira14.pdf

Orgut, I. S., Ivy, J., Uzsoy, R., & Wilson, J. R. (2016a). Modeling for the equitable and effective distribution of donated food under capacity constraints. *IIE Transactions*, *48*(3), 252-266. https://doi.org/10.1080/0740817X.2015.1063792

Orgut, I. S., Brock III, L. G., Davis, L. B., Ivy, J. S., Jiang, S., Morgan, S. D., Uzsoy, R., Hale, C., & Middleton, E. (2016b). Achieving equity, effectiveness, and efficiency in food bank operations: Strategies for feeding America with implications for global hunger relief. In *Advances in managing humanitarian operations* (pp. 229-256). Springer, Cham. https://doi.org/10.1007/978-3-319-24418-1_11

Orgut, I. S., Ivy, J., & Uzsoy, R. (2017). Modeling for the equitable and effective distribution of food donations under stochastic receiving capacities. *IISE Transactions*, *49*(6), 567-578. https://doi-org.ezproxy.lib.purdue.edu/10.1080/24725854.2017.1300358

Orgut, I. S., Ivy, J. S., Uzsoy, R., & Hale, C. (2018). Robust optimization approaches for the equitable and effective distribution of donated food. *European Journal of Operational Research*, *269*(2), 516-531. https://doi.org/10.1016/J.EJOR.2018.02.017

Ortuño, J. C., & Gómez Padilla, A. (2017). Assembly of customized food pantries in a food bank by fuzzy optimization. *Journal of Industrial Engineering and Management (JIEM)*, *10*(4), 663-686. https://doi.org/10.3926/jiem.2160

Pahl, G., & Beitz, W. (2013). *Engineering design: a systematic approach*. Springer Science & Business Media. https://books.google.com/books?hl=en&lr=&id=4uvSBwAAQBAJ&oi=fnd&pg=PR12&dq=pahl+beitz&ots=pmnkwdkdSn&sig=DhXzVdfoZMfYgfkdX6ZC_DnJy3k#v=onepage&q=pahl%20beitz&f=false

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497-505. https://doi.org/10.1016/j.omega.2004.07.024

Paul, S., and Davis, L. B. (2019). Predicting Food Donor Contribution Behavior Using Support Vector Regression. In *Proceedings of the 2019 IISE Annual Conference, Orlando, FL, USA*. https://par.nsf.gov/servlets/purl/10107757

Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a service: why, what, how, and what research is needed?. *Journal of service research*, *12*(2), 175-189. https://doi.org/10.1177/1094670509344213

Parfitt, J., Barthel, M., & Macnaughton, S. (2010). Food waste within food supply chains: quantification and potential for change to 2050. *Philosophical transactions of the royal society B: biological sciences*, *365*(1554), 3065-3081. https://doi.org/10.1098/rstb.2010.0126

Parmenter, D. (2015). *Key performance indicators: developing, implementing, and using winning KPIs*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=bKkxBwAAQBAJ&oi=fnd&pg=PA101&dq=D.+Parmenter,+Key+Performance+Indicators+(KPI):+Developing,+Implementing,+and+Using+Winning+KPIs,+Third+Ed.,+John+Wiley+%26+Sons,+Inc.,+New&ots=cZT_jZe0bo&sig=Zao05UcBhbsVVbA45B7l3ZkHq-0#v=onepage&q&f=false

Phadke, R., Prasad, V., Nagaraj, H. C., & Bhograj, A. (2020). Univariate data-driven models for glucose level prediction of CGM sensor dataset for T1DM management. *Sādhanā*, *45*(1), 1-17. https://doi.org/10.1007/s12046-020-1277-8S

Phillips, C., Hoenigman, R., Higbee, B., & Reed, T. (2013). Understanding the sustainability of retail food recovery. *PLOS one*, *8*(10), e75530. https://doi.org/10.1371/journal.pone.0075530

Pollastri, A., & Maffenini, W. (2018). Estimation of the Percentage of Aid Received from the Food Bank Using a Stratified Sampling. *Social Indicators Research*, *136*(1), 41-49. https://doi.org/10.1007/s11205-016-1514-z

Prendergast, C. (2016). The Allocation of Food to Food Banks. *EAI Endorsed Trans. Serious Games*, *3*(10), e4. https://ipl.econ.duke.edu/seminars/system/files/seminars/1491.pdf

Prendergast, C. (2017). How food banks use markets to feed the poor. *Journal of Economic Perspectives*, *31*(4), 145-62. https://doi.org/10.1257/jep.31.4.145

Raffensperger, J. F., Girardini, J., & Mastbaum, J. (2020). *New Metrics and Visualizations to Help the Army Reduce Customer Wait Time*. RAND Santa Monica United States. https://apps.dtic.mil/sti/pdfs/AD1090524.pdf

Reihaneh, M., & Ghoniem, A. (2018). A multi-start optimization-based heuristic for a food bank distribution problem. *Journal of the operational research society*, *69*(5), 691-706. https://doi.org/10.1057/s41274-017-0220-9

Rey, D., Almi'ani, K., & Nair, D. J. (2018). Exact and heuristic algorithms for finding envy-free allocations in food rescue pickup and delivery logistics. *Transportation Research Part E: Logistics and Transportation Review*, *112*, 19-46. https://doi.org/10.1016/j.tre.2018.02.001

Riches, G. (2002). Food banks and food security: welfare reform, human rights and social policy. Lessons from Canada?. *Social Policy & Administration*, *36*(6), 648-663. https://doi.org/10.1111/1467-9515.00309

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Russell, T. E. (2005). *The humanitarian relief supply chain: analysis of the 2004 South East Asia earthquake and tsunami* (Doctoral dissertation, Massachusetts Institute of Technology). https://dspace.mit.edu/handle/1721.1/33352

Schleich, B. R., Lam, S. S., Yoon, S. W., Tajik, W., Goldstein, M. J., & Irving, H. (2013). A Neural Network-based Approach for Predicting Organ Donation Potential. In *IIE Annual Conference. Proceedings* (p. 1532). Institute of Industrial and Systems Engineers (IISE). https://search.proquest.com/docview/1471958899/fulltextPDF/3E60E99B57494A2EPQ/1?accountid=13360

Schnaubelt, M. (2019). *A comparison of machine learning model validation schemes for non-stationary time series data* (No. 11/2019). FAU Discussion Papers in Economics. http://hdl.handle.net/10419/209136

Schneider, K., & Nurre, S. G. (2019). A multi-criteria vehicle routing approach to improve the compliance audit schedule for food banks. *Omega*, *84*, 127-140. https://doi.org/10.1016/j.omega.2018.05.002

Shi, Z. R., Yuan, Y., Lo, K., Lizarondo, L., & Fang, F. (2020, April). Improving efficiency of volunteer-based food rescue operations. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 08, pp. 13369-13375). https://doi.org/10.1609/aaai.v34i08.7051

Shim, J. K., Siegel, J. G., & Shim, A. I. (2011). *Budgeting basics and beyond* (Vol. 574). John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=FsioDgAAQBAJ&oi=fnd&pg=PR11&dq=Glossary+of+Budgeting+and+Planning+Terms+shim&ots=qQzenf05rl&sig=eae_TaGjd0nKz1Hq84g-5NimEKw#v=onepage&q=Glossary%20of%20Budgeting%20and%20Planning%20Terms%20shim&f=false

Shrestha, A. R. (2009). Feed my people food bank projecta case study of space planning and costing of warehouse. https://minds.wisconsin.edu/bitstream/handle/1793/43329/2009shresthaa.pdf?sequence=1

Solak, S., Scherrer, C., & Ghoniem, A. (2014). The stop-and-drop problem in nonprofit food distribution networks. *Annals of Operations Research*, *221*(1), 407-426. https://doi.org/10.1007/s10479-012-1068-7

Sönmez, E., Lee, D., Gómez, M. I., & Fan, X. (2016). Improving food bank gleaning operations: An application in New York state. *American Journal of Agricultural Economics*, *98*(2), 549-563. https://econpapers.repec.org/scripts/redir.pf?u=http%3A%2F%2Fhdl.handle.net%2F10.1093%2Fajae%2Faav069;h=repec:oup:ajagec:v:98:y:2016:i:2:p:549-563

Srivastav, A., Tewari, A., & Dong, B. (2013). Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. *Energy and Buildings*, *65*, 438-447. https://doi.org/10.1016/j.enbuild.2013.05.037

Stirrup, J. (2016). *Tableau dashboard cookbook*. Packt Publishing Ltd. https://books.google.com/books?hl=en&lr=&id=lamkAwAAQBAJ&oi=fnd&pg=PP1&dq=Stirrup,+Jen.+2014.+Tableau+Dashboard+Cookbook+Over+40+Recipes+on+Designing+Professional+Dashboards+by+Implementing+Data+Visualization+Principles.+www.packtpub.com.&ots=2JitpDqfiQ&sig=1ofSBrqJMtAp20W21tLxwYNON_4#v=onepage&q&f=false

Sucharitha, R. S., & Lee, S. (2018). New Policy Design for Food Accessibility to the People in Need. *arXiv preprint arXiv:1909.08648*

Sucharitha, R. S., & Lee, S. (2019). Application of Clustering Analysis for Investigation of Food Accessibility. *Procedia Manufacturing*, *39*, 1809-1816. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.promfg.2020.01.258

Taneja, K. (2020). *Machine Learning Based Data Driven Modelling of Time Series of Power Plant Data* (Master's thesis, University of Waterloo). http://hdl.handle.net/10012/15792

Tarasuk, V., St-Germain, A. A. F., & Loopstra, R. (2020). The relationship between food banks and food insecurity: insights from Canada. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *31*(5), 841-852. https://doi.org/10.1007/s11266-019-00092-w

Tümer, A. E., & Akkuş, A. (2018). Forecasting gross domestic product per capita using artificial neural networks with non-economical parameters. *Physica A: Statistical Mechanics and its Applications*, *512*, 468-473. https://doi.org/10.1016/j.physa.2018.08.047

USDA ERS - Interactive Charts and Highlights. (n.d.). USDA. Retrieved June 25, 2018, from https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/interactive-charts-and-highlights/#trends

Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, 281-287. https://books.google.com/books?hl=en&lr=&id=QpD7n95ozWUC&oi=fnd&pg=PA281&dq=Support+Vector+Method+for+Function+Approximation,+Regression+Estimation,+and+Signal+Processing%C2%B7&ots=iEktsGSZgD&sig=45uyFPvD7vl5y1JjMlfVb8SxU0Q#v=onepage&q=Support%20Vector%20Method%20for%20Function%20Approximation%2C%20Regression%20Estimation%2C%20and%20Signal%20Processing%C2%B7&f=false

Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, N. (2012). A comparative study of various clustering algorithms in data mining. *International Journal of Engineering Research and Applications (IJERA)*, *2*(3), 1379-1384. https://d1wqtxts1xzle7.cloudfront.net/28319731/ID2313791384.pdf?1348378439=&response-content-disposition=inline%3B+filename%3DIJERA_www_ijera_com.pdf&Expires=1615845481&Signature=TiqnJQDrtH58uMq53KRRs-3nkgUh0DjFNv2gBUbSDt6cFc-eBVbkUy6wjJC19FBEBk7fKJbY5Zhb8c4Mftqgu61TiA4m6AcJHjh1FBEom6f7G~NbZDvqkBxrmAaLvmjvbNHbNFskEnLOAVxytnQtYC8lFoxFnA9wApUEufEKGVMNJ0vcD6SL64q~XQMLMoC9~wjv8BGK8xtv6m4EUFVs-siYA3VuQpCetT80b32~sGAG-iYX6G9aW7JkAuFSD4StRTR6N0bo73p~OUHp~tk2f6H4sSIkSdCAPhQbLdZlaVta0jqx5udLVDTnt5Sz90sYfofuso2NJHSRROjyBMLdrA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Vilarinho, S., Lopes, I., & Sousa, S. (2018). Developing dashboards for SMEs to improve performance of productive equipment and processes. *Journal of Industrial Information Integration*, *12,* 13-22. https://doi.org/10.1016/j.jii.2018.02.003

Vozoris, N. T., & Tarasuk, V. S. (2003). Household food insufficiency is associated with poorer health. *The Journal of nutrition*, *133*(1), 120-126. https://doi.org/10.1093/jn/133.1.120

Waity, J. F. (2016). Spatial inequality in access to food assistance in Indiana. *Sociological Inquiry*, *86*(1), 103-126. https://doi.org/10.1111/soin.12098

Wang, Z., Da Cunha, C., Ritou, M., & Furet, B. (2019). Comparison of K-means and GMM methods for contextual clustering in HSM. *Procedia Manufacturing*, *28*, 154-159. https://doi.org/10.1016/j.promfg.2018.12.025

Wetherill, M. S., White, K. C., & Seligman, H. K. (2019). Nutrition-focused food banking in the United States: a qualitative study of healthy food distribution initiatives. *Journal of the Academy of Nutrition and Dietetics*, *119*(10), 1653-1665. https://doi-org.ezproxy.lib.purdue.edu/10.1016/j.jand.2019.04.023

Wittman, H., Chappell, M. J., Abson, D. J., Kerr, R. B., Blesh, J., Hanspach, J., ... & Fischer, J. (2017). A social–ecological perspective on harmonizing food security and biodiversity conservation. *Regional Environmental Change*, *17*(5), 1291-1301. https://doi.org/10.1007/s10113-016-1045-9

Wongso, E., Nateghi, R., Zaitchik, B., Quiring, S., & Kumar, R. (2019). Leveraging statistical learning theory to characterize the US water consumption. https://www.essoar.org/doi/abs/10.1002/essoar.10500810.1

Wu, E., Villani, J., Davis, A., Fareed, N., Harris, D. R., Huerta, T. R., ... & Oga, E. A. (2020). Community dashboards to support data-informed decision-making in the HEALing Communities Study. *Drug and alcohol dependence*, *217*, 108331. https://doi.org/10.1016/j.drugalcdep.2020.108331

Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, *2*(2), 165-193. https://doi.org/10.1007/s40745-015-0040-1

Yang, Q. (2018). Optimal Allocation Algorithm for Sequential Resource Allocation in the Context of Food Banks Operations. (Doctoral dissertation, Cornell University). https://doi.org/10.7298/X4R78CGC

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41-59. https://doi.org/10.1016/j.accinf.2011.08.002

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega*, *41*(3), 517-524. https://doi.org/10.1016/j.omega.2012.06.005

9. APPENDIX

9.1 Data description (Information Classes)

Table 9.1 Training datasets and corresponding ACF, PACF plots for each information class

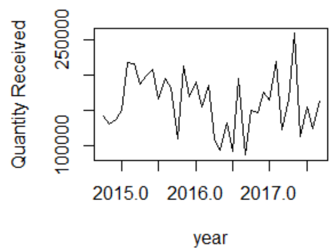
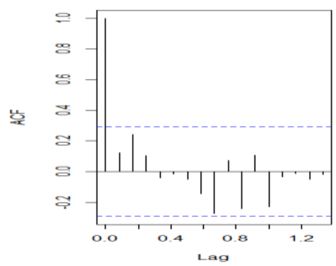
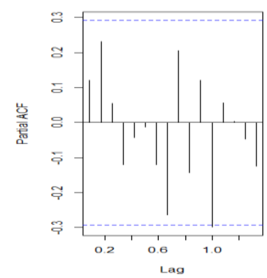
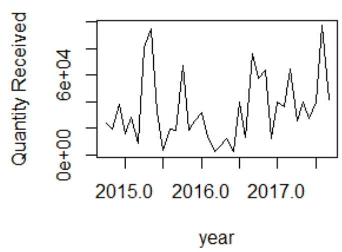
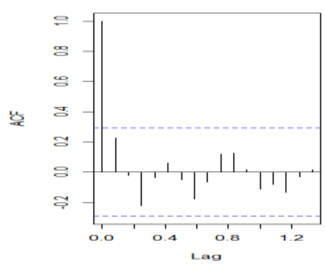
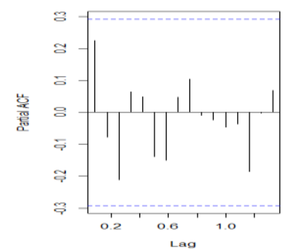
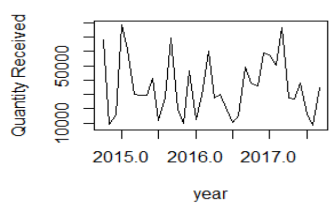
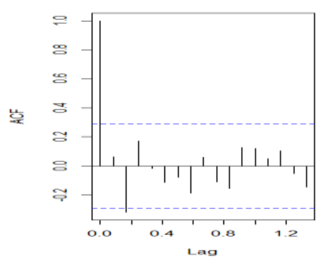
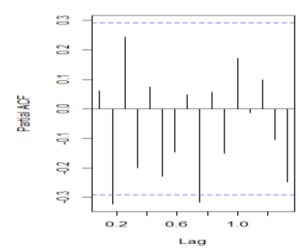
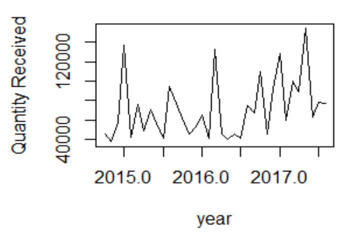
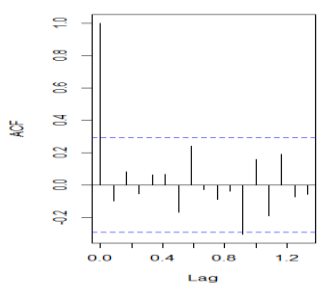
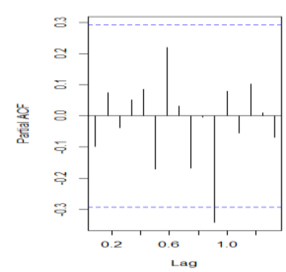
<table>
<tr><th>Information class</th><th>Training data</th><th>ACF Plot</th><th>PACF Plot</th></tr>
<tr><td>C_{FB}</td><td></td><td></td><td></td></tr>
<tr><td>C_{FC}</td><td></td><td></td><td></td></tr>
<tr><td>C_{FD}</td><td></td><td></td><td></td></tr>
<tr><td>C_{FM}</td><td></td><td></td><td></td></tr>
</table>

Table 9.1 continued

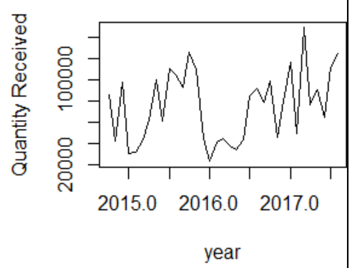
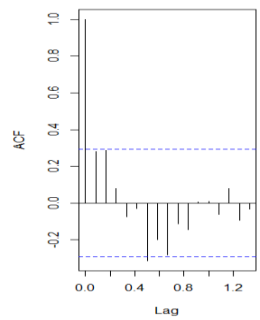
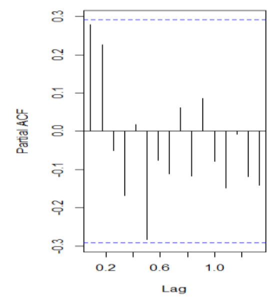
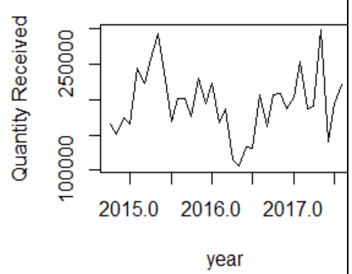
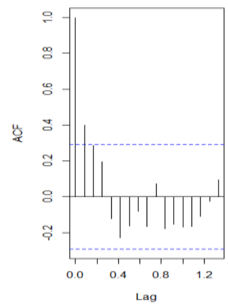
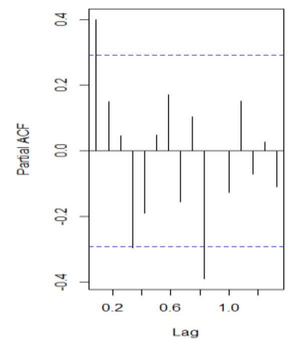
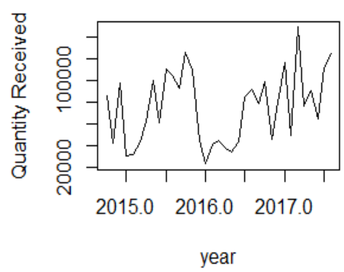
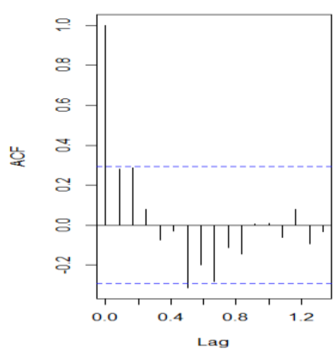
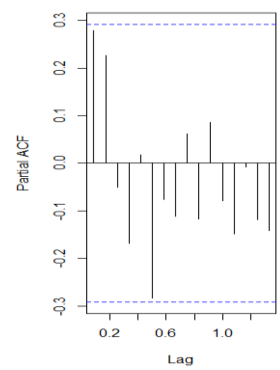
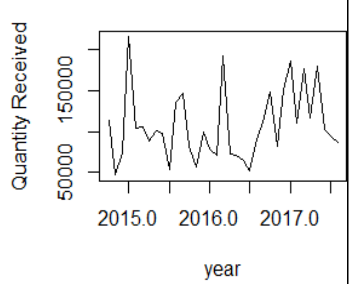
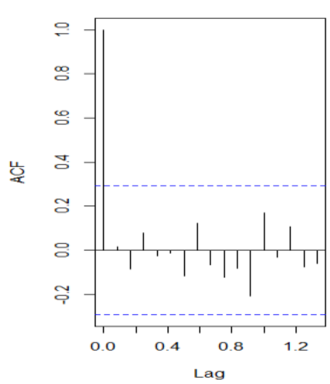
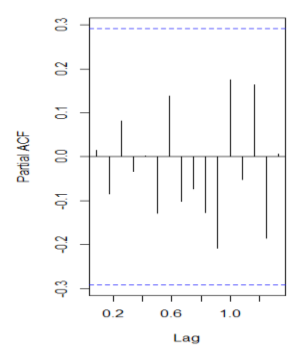
C_{FF}			
C_{SD}			
C_{SF}			
C_{SFZ}			

Table 9.1 continued

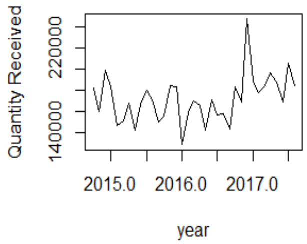
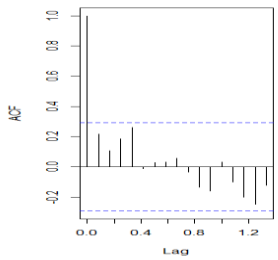
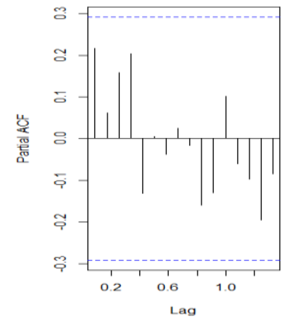
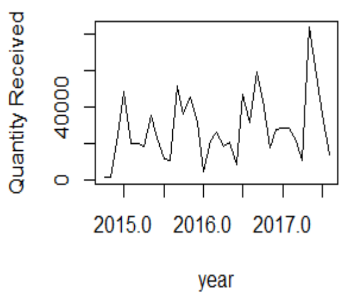
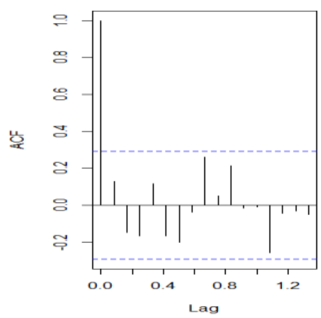
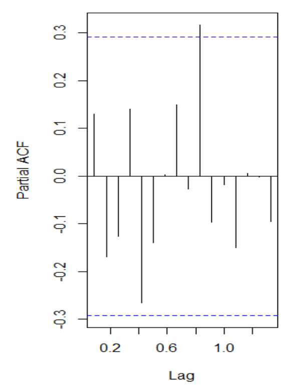
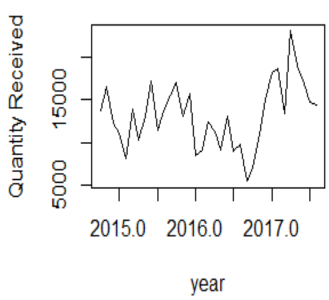
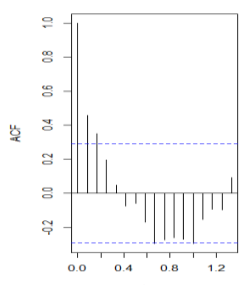
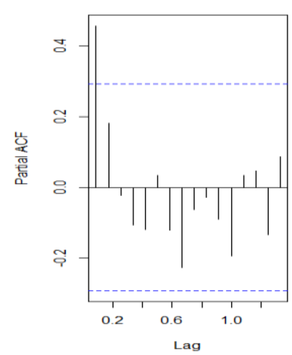
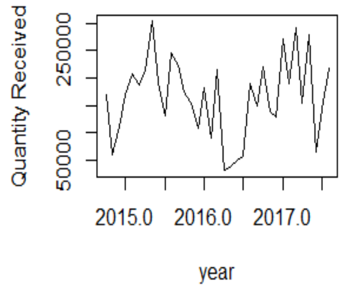
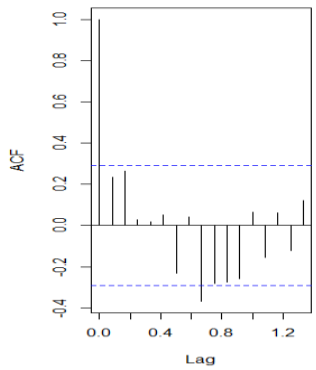
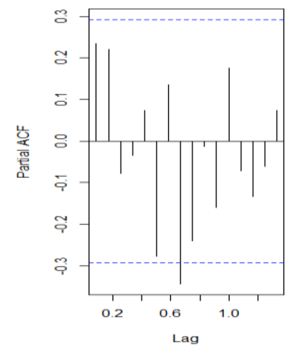
C_{DR}			
C_{DM}			
C_{DRT}			

Table 9.1 continued

C_{DO}			

9.2 Cross-Validation for time series

There are two methods of walk-forward cross-validation of time series: sliding window and expanding window (Schnaubelt, 2019).

In sliding window, we have the same training size, but the window size is slid across the data to create multiple train-test pairs. In expanding window, we expand the training size from a particular starting size to a maximum size.

From the results, we see that by using expanding window cross-validation, we observe better performance results than implementing sliding window cross-validation. Considering the three-year monthly datasets, expanding window provides a good balance between creating enough pairs while maximizing the dataset size in hand (Bergmeir et al., 2015). Section 9.2.1 provides are the performance results of the remaining information classes with the aggregated dataset results being provided in Chapter 3 for expanding window cross validation study. Section 9.2.2 provides the performance results of all the information classes considering the sliding window cross validation. Finally, Section 9.2.3 provides the forecast plots for the remaining information classes with the actual test values with a 6-months ahead forecast depicted in Fig 9.1 – 9.12. We provide the forecast plots based on the expanding window cross validation results based on better predictive performance of this method than sliding window cross validation method. From observing the forecasts, we see that the forecasts of the best models for each dataset appear to be good at predicting the general direction of the food supply.

9.2.1 Expanding window cross-validation results

Table 9.2 Quantitative measures of performance for different forecasting models for Food type (Breads and Cereals) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	37629.27	32299.75	21.84
SVM	**Gamma = 0.01, cost = 0.1**	**31228.48**	**26737.36**	**18.84**
NNAR	(1,1,2)[12]	35754.68	30260.33	21.74
Hybrid ARIMA-SVM		35755.06	29928.45	21.67
Hybrid ARIMA-NNAR		35857.23	30608.03	22.19

Table 9.3 Quantitative measures of performance for different forecasting models for Food type (Confectionaries) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	25431.54	20561.05	153.60
SVM	Gamma = 0.03, cost = 0.9	25639.61	20464.66	158.20
NNAR	(1,1,2)[12]	22712.13	18249.76	144.42
Hybrid ARIMA-SVM		14138.32	11324.64	90.69
Hybrid ARIMA-NNAR		**13877.69**	**11079.16**	**94.53**

Table 9.4 Quantitative measures of performance for different forecasting models for Food type (Dairy products) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,1)	17687.82	14962.06	68.57
SVM	Gamma = 0.01, cost = 0.1	15805.57	13100.79	59.30
NNAR	(3,1,2)[12]	14984.2	12833.24	62.45
Hybrid ARIMA-SVM		13598.64	11386.53	48.34
Hybrid ARIMA-NNAR		**13196.26**	**11243.16**	**52.15**

Table 9.5 Quantitative measures of performance for different forecasting models for Food type (Fruits and Vegetables) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	30982.93	27046.72	47.92
SVM	Gamma = 0.1, cost = 0.8	26988.69	22209.22	38.19
NNAR	(1,1,2)[12]	26973.69	22462.75	39.42
Hybrid ARIMA-SVM		**25587.94**	**21187.33**	**36.59**
Hybrid ARIMA-NNAR		26756.82	22730.61	39.23

Table 9.6 Quantitative measures of performance for different forecasting models for Food type (Meats) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	30006.96	23341.9	35.73
SVM	Gamma = 0.1, cost = 0.7	21165.84	15819.86	22.47
NNAR	**(1,1,2)[12]**	**20397.1**	**15416.19**	**22.83**
Hybrid ARIMA-SVM		23143.59	17816.82	26.09
Hybrid ARIMA-NNAR		26007.44	20458.7	30.52

Table 9.7 Quantitative measures of performance for different forecasting models for Storage type (Dry) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,0)	41310.32	32536.63	17.61
SVM	Gamma = 0.1, cost = 0.7	37729.82	29785.6	16.22
NNAR	(13,1,7)[12]	22218.32	17011.39	9.09
Hybrid ARIMA-SVM		35962.55	28439.12	15.50
Hybrid ARIMA-NNAR		**21662.17**	**17129.73**	**9.59**

Table 9.8 Quantitative measures of performance for different forecasting models for Storage type (Fresh) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	30982.93	27046.72	47.92
SVM	Gamma = 0.07, cost = 0.1	26988.69	22209.22	38.19
NNAR	(1,1,2)[12]	26973.69	22462.75	39.42
Hybrid ARIMA-SVM		26756.82	22730.61	39.23
Hybrid ARIMA-NNAR		**25587.94**	**21187.33**	**36.59**

Table 9.9 Quantitative measures of performance for different forecasting models for Storage type (Frozen) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	42645.53	32025.29	33.91
SVM	Gamma = 0.01, cost = 0.1	30164.57	23418.99	23.12
NNAR	**(3,1,2)[12]**	**22823.26**	**18096.11**	**19.02**
Hybrid ARIMA-SVM		33497.63	26179.13	26.91
Hybrid ARIMA-NNAR		29004.3	23099.74	24.36

Table 9.10 Quantitative measures of performance for different forecasting models for Donor type (Manufacturers) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(0,0,0)	17611.47	13804.56	227.85
SVM	Gamma = 0.07, cost = 1	15024.74	11378.1	197.06
NNAR	(2,1,2)[12]	15284.4	10983.48	170.76
Hybrid ARIMA-SVM		**14773.63**	**11203.86**	**148.21**
Hybrid ARIMA-NNAR		15567.72	11658.01	148.59

Table 9.11 Quantitative measures of performance for different forecasting models for Donor type (others) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,0)	65984.29	54924.25	54.35
SVM	Gamma = 0.01, cost = 0.1	52018.07	40403.17	42.81
NNAR	(2,1,2)[12]	36564.54	28713.51	34.91
Hybrid ARIMA-SVM		51409.15	39521.66	39.37
Hybrid ARIMA-NNAR		**33568.09**	**28133.26**	**30.48**

Table 9.12 Quantitative measures of performance for different forecasting models for Donor type (Restaurants) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)(0,0,1)[12]	2989.62	2457.64	21.18
SVM	Gamma = 0.09, cost = 0.4	3001.57	2548.24	20.47
NNAR	(2,1,2)[12]	2963.87	2448.80	20.86
Hybrid ARIMA-SVM		2846.76	2263.08	19.95
Hybrid ARIMA-NNAR		**2713.37**	**2288.54**	**19.03**

Table 9.13 Quantitative measures of performance for different forecasting models for Donor type (Retailers) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(0,0,0)	21289.28	16572.77	9.80
SVM	Gamma = 0.1, cost = 1	16921.58	12602.88	7.52
NNAR	(3,1,2)[12]	17410.24	12059.98	7.13
Hybrid ARIMA-SVM		**16363.95**	**12739.06**	**7.56**
Hybrid ARIMA-NNAR		17518.12	13081.24	7.71

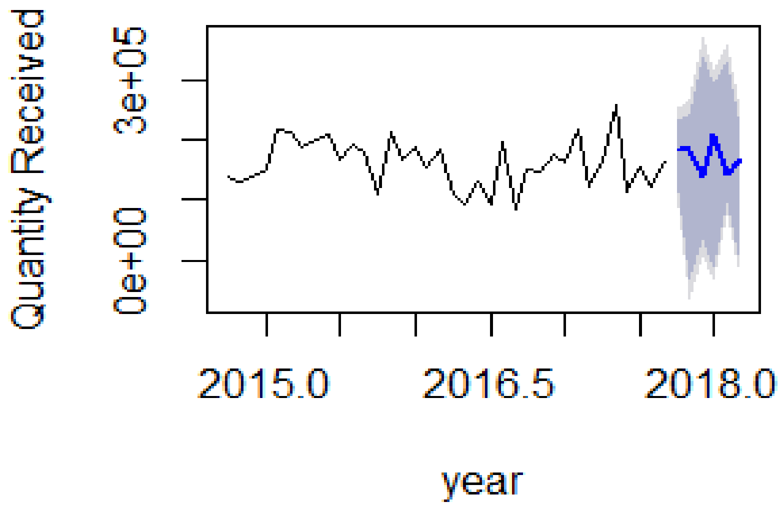


Figure 9.1 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Food type (Bread and Cereals) dataset

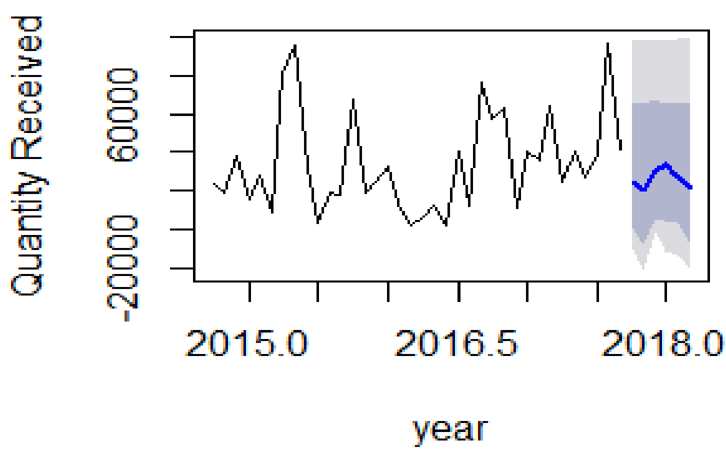


Figure 9.2 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Food type (Confectionaries) dataset

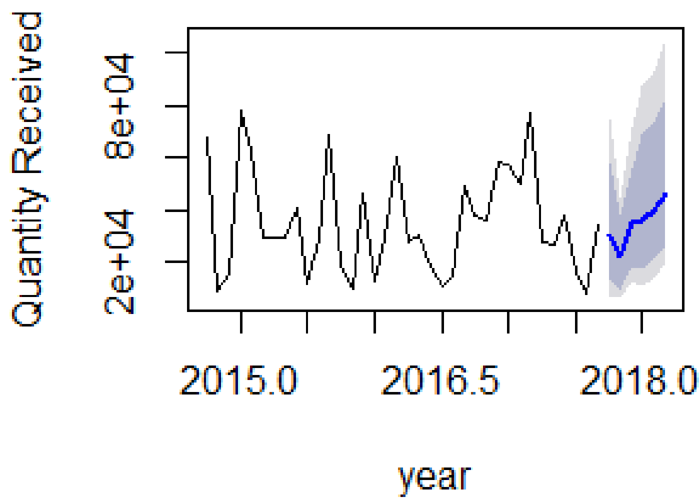


Figure 9.3 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Food type (Dairy products) data set

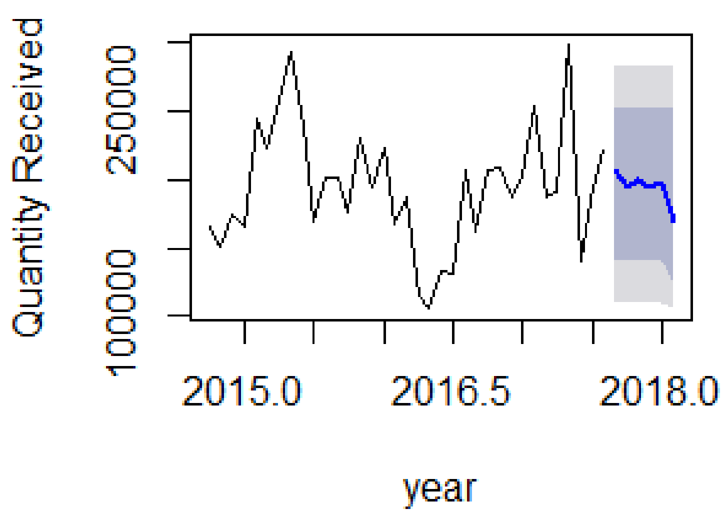


Figure 9.4 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Storage type (Dry) data set

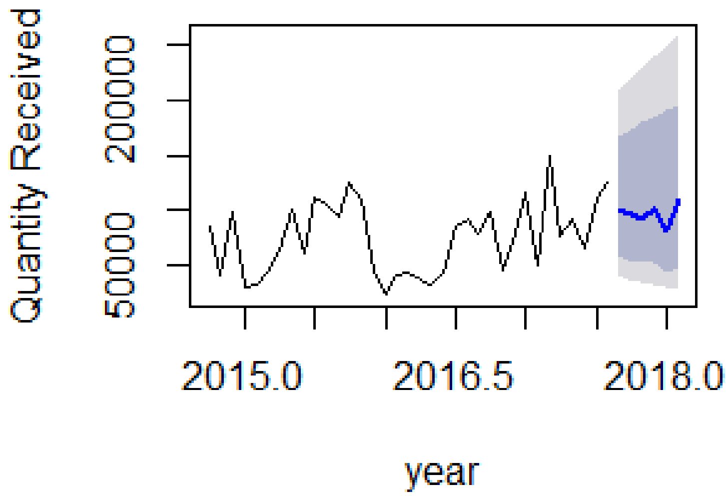


Figure 9.5 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Storage type (Fresh) data set

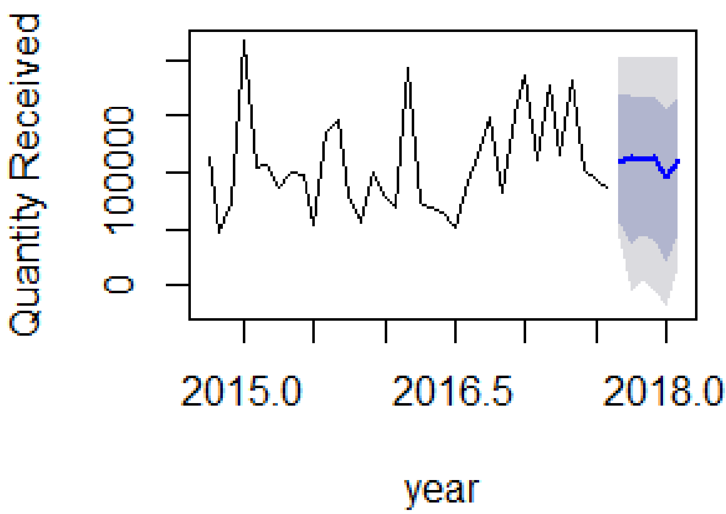


Figure 9.6 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Storage type (Frozen) data set

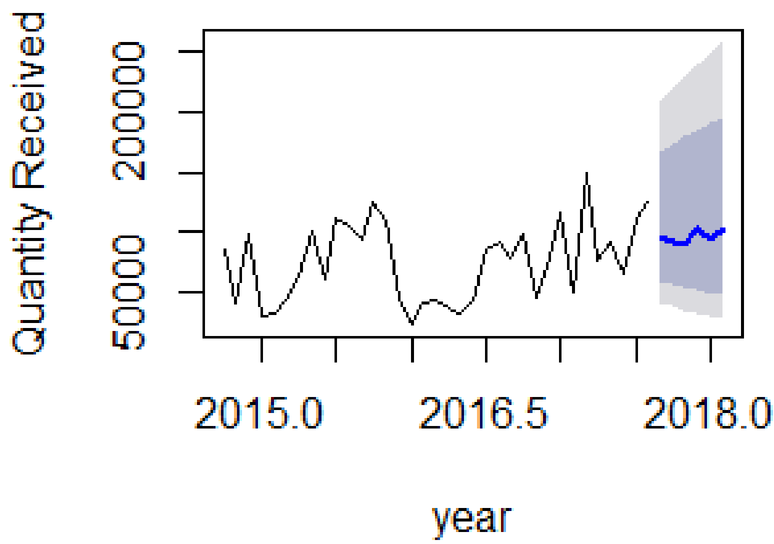


Figure 9.7 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Food type (Fruits and Vegetables) data set

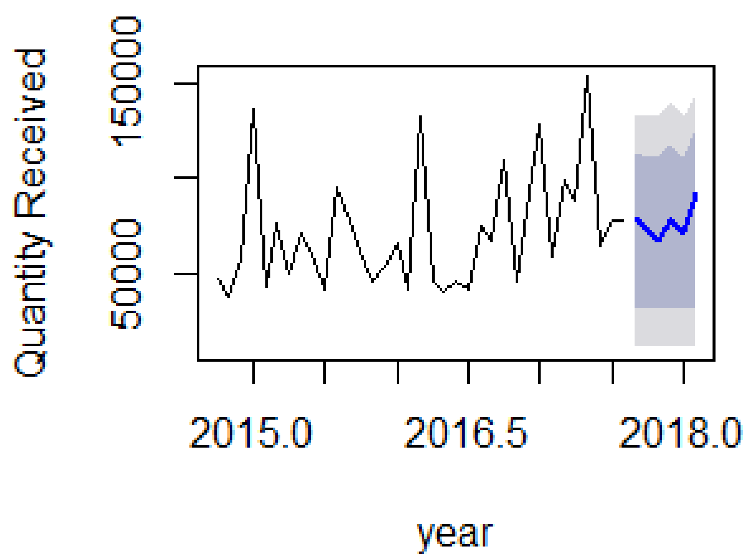


Figure 9.8 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Food type (Meats) data set

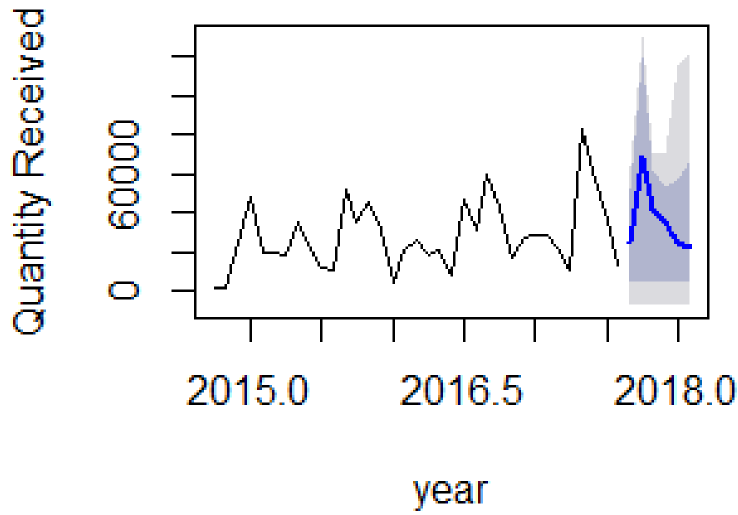


Figure 9.9 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Donor type (Manufacturers) data set

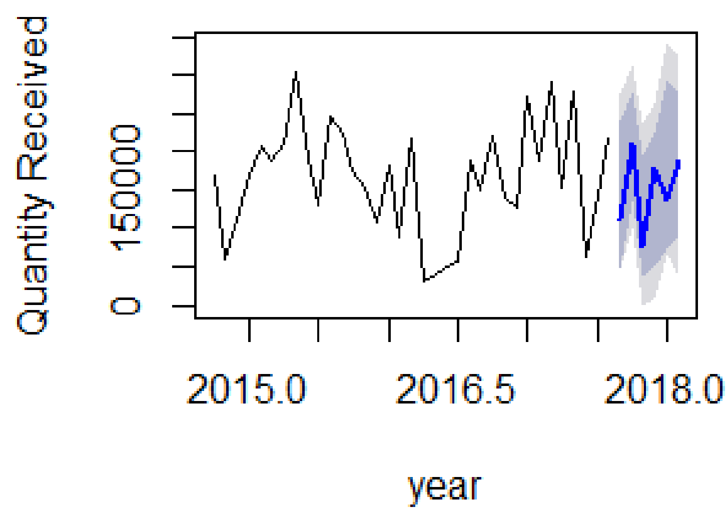


Figure 9.10 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Donor type (Others) data set

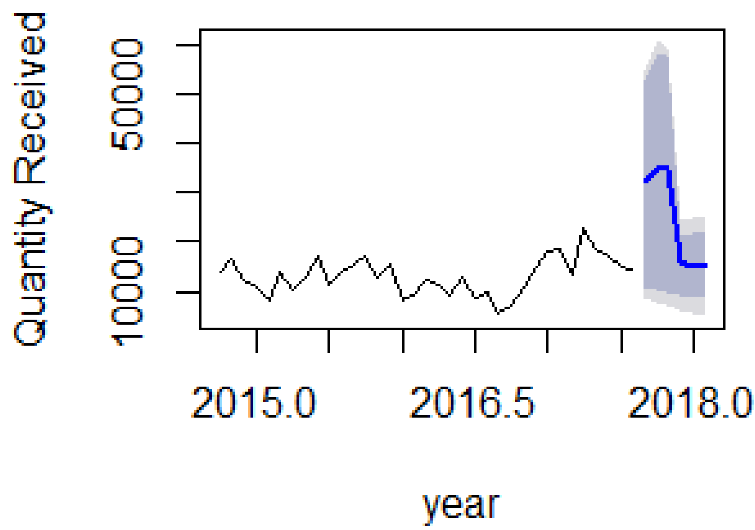


Figure 9.11 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Donor type (Restaurants) data set

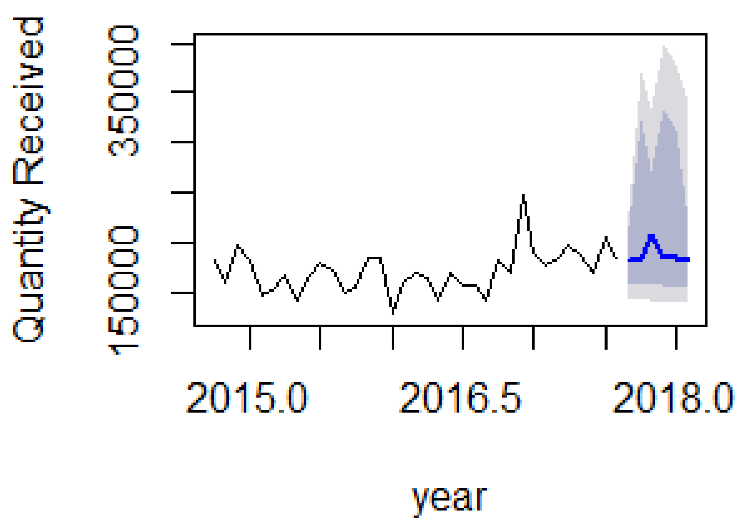


Figure 9.12 Actual vs predicted forecasts (using ARIMA-NNAR model) of the Donor type (Retailers) data set

9.2.2 Sliding window cross-validation results

Table 9.14 Quantitative measures of performance for different forecasting models for aggregated dataset

	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,1)	79665.3	65743.31	18.71
SVM	Gamma = 0.03, cost = 1	65931.71	54246.26	15.68
NNAR	(9,1,6)[12]	62525.96	50598.75	14.68
Hybrid ARIMA-SVM		43464.33	36096.15	10.38
Hybrid ARIMA-NNAR		**36082.95**	**28563.39**	**8.28**

Table 9.15 Quantitative measures of performance for different forecasting models for Food type (Breads and Cereals) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	37213.81	32113.54	21.82
SVM	**Gamma = 0.01, cost = 0.1**	**31237.45**	**27253.59**	**20.36**
NNAR	(1,1,2)[12]	36171.06	30848.47	21.75
Hybrid ARIMA-SVM		35303.96	29365.23	20.54
Hybrid ARIMA-NNAR		37289.04	31098.45	22.14

Table 9.16 Quantitative measures of performance for different forecasting models for Food type (Confectionaries) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	26609.33	19504.84	133.13
SVM	Gamma = 0.03, cost = 0.9	25574.55	19558.02	155.94
NNAR	(1,1,2)[12]	23983.94	19204.1	140.81
Hybrid ARIMA-SVM		20479.97	16479.01	160.80
Hybrid ARIMA-NNAR		**17668.32**	**13902.4**	**127.69**

Table 9.17 Quantitative measures of performance for different forecasting models for Food type (Dairy products) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,1)	20068.73	16683.95	75.53
SVM	Gamma = 0.01, cost = 0.1	17758.89	14980.87	68.12
NNAR	(3,1,2)[12]	16387.97	13377.62	55.21
Hybrid ARIMA-SVM		14870.39	11362.67	39.58
Hybrid ARIMA-NNAR		**14471.76**	**10925.86**	**42.87**

Table 9.18 Quantitative measures of performance for different forecasting models for Food type (Fruits and Vegetables) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	32873.65	26895.97	45.44
SVM	Gamma = 0.1, cost = 0.8	32777.15	26697.97	43.31
NNAR	(1,1,2)[12]	32779.11	26686.82	44.25
Hybrid ARIMA-SVM		**26824.17**	**21245.31**	**35.31**
Hybrid ARIMA-NNAR		31574.9	27677.43	48.35

Table 9.19 Quantitative measures of performance for different forecasting models for Food type (Meats) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	31712.77	22935.76	29.94
SVM	Gamma = 0.1, cost = 0.7	26019.78	19497.68	26.07
NNAR	**(1,1,2)[12]**	**23696.1**	**17817.82**	**25.04**
Hybrid ARIMA-SVM		25819.2	20663.51	30.05
Hybrid ARIMA-NNAR		30635.35	23203.13	32.87

Table 9.20 Quantitative measures of performance for different forecasting models for Storage type (Dry) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,0)	44219.97	33014.33	17.60
SVM	Gamma = 0.1, cost = 0.7	41063.64	31789.52	17.55
NNAR	(13,1,7)[12]	27896.69	21698.09	13.09
Hybrid ARIMA-SVM		36920.36	29058.88	16.09
Hybrid ARIMA-NNAR		**27476.55**	**22082.37**	**12.87**

Table 9.21 Quantitative measures of performance for different forecasting models for Storage type (Fresh) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)	32873.65	26895.97	45.44
SVM	Gamma = 0.07, cost = 0.1	32779.11	26686.82	44.25
NNAR	(1,1,2)[12]	32777.15	26697.37	43.31
Hybrid ARIMA-SVM		31574.9	27677.43	48.35
Hybrid ARIMA-NNAR		**26824.17**	**21245.31**	**35.31**

Table 9.22 Quantitative measures of performance for different forecasting models for Storage type (Frozen) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,1)	44565.91	33001.2	31.02
SVM	Gamma = 0.01, cost = 0.1	34196.99	25803.37	23.85
NNAR	**(3,1,2)[12]**	**30269.37**	**22191.3**	**20.91**
Hybrid ARIMA-SVM		38189.51	30081.59	28.06
Hybrid ARIMA-NNAR		33377.54	25954.64	25.85

Table 9.23 Quantitative measures of performance for different forecasting models for Donor type (Manufacturers) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(0,0,0)	18619.19	13606.42	138.86
SVM	Gamma = 0.07, cost = 1	17806.53	13856.17	178.28
NNAR	(2,1,2)[12]	17807.43	13857.21	178.33
Hybrid ARIMA-SVM		**14974.97**	**11383.95**	**128.05**
Hybrid ARIMA-NNAR		17969.39	13496.83	157.65

Table 9.24 Quantitative measures of performance for different forecasting models for Donor type (others) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(2,0,0)	72133.3	58500.32	64.18
SVM	Gamma = 0.01, cost = 0.1	53876.31	41406.46	44.64
NNAR	(2,1,2)[12]	48702.25	42219.33	43.34
Hybrid ARIMA-SVM		53031.36	41820.46	44.64
Hybrid ARIMA-NNAR		**37948.26**	**31026.77**	**39.57**

Table 9.25 Quantitative measures of performance for different forecasting models for Donor type (Restaurants) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(1,0,0)(0,0,1)[12]	3102.39	2624.04	21.34
SVM	Gamma = 0.09, cost = 0.4	3227.67	2647.87	22.96
NNAR	(2,1,2)[12]	3028.97	2409.0	20.82
Hybrid ARIMA-SVM		3001.64	2443.6	19.83
Hybrid ARIMA-NNAR		**2973.12**	**2242.98**	**19.96**

Table 9.26 Quantitative measures of performance for different forecasting models for Donor type (Retailers) dataset

Model	Parameters	RMSE	MAE	MAPE
ARIMA	(0,0,0)	21781.07	15577.96	9.00
SVM	Gamma = 0.1, cost = 1	21610.14	15306.77	8.90
NNAR	(3,1,2)[12]	17057.29	11796.74	6.85
Hybrid ARIMA-SVM		**16858.11**	**12614.46**	**7.31**
Hybrid ARIMA-NNAR		21629.79	14990.05	8.76

VITA

RAHUL SRINIVAS SUCHARITHA

EDUCATION

Purdue University, West Lafayette, IN

Fall 2020

Ph.D. Candidate, Industrial Engineering, **GPA 3.72/4.0**

- **Ph.D. Thesis:** Logistics of Food Bank Operations in Distribution of Donated Foods
- Minor in Statistics

Purdue University, West Lafayette, IN

Spring 2015

Master of Science, Industrial Engineering, **GPA 3.85/4.0**

College of Engineering Guindy, Chennai, India

Spring 2013

Bachelor of Engineering in Mechanical Engineering, **GPA 8.84/10.0**

- **Bachelors Thesis:** Design and Fabrication of hand-driven tricycle-cum-wheelchair for the physically challenged

PROFESSIONAL EXPERIENCE

Greif, Inc., Delaware, OH

Summer 2020

Supply chain Intern

- Performed Root Cause Analysis to investigate the changes in performance metrics and to integrate proposed data visualization framework with the organization's strategic goals and operations
- Managed designing and execution of custom reporting and Tableau dashboard solutions while coordinating work across multi-departmental resources and communicating progress to internal and external stakeholders
- Developed design documentation procedures to enable Business Intelligence team for customization of Tableau dashboard designs ensuring continuous improvement and stakeholder satisfaction

Oscar Winski Company Inc., Lafayette, IN

Summer 2017

Quality and Logistics Intern

- Developed Quality inspection documents as an inclusion to the PPAP (Production Part Approval Process) requirements for all the machines in the plant by the utilization of 5S methodology

- Developed multiple Process Maps and SIPOC (Supply Input Process Output) for four different stages (Driver Check-in, Loading, Unloading, and Driver Check-out) in the current process flow of Fleet and Driver logistics

- Recommended several changes in work processes and driver treatment in the plant accomplishing a reduction in fleet-time spent on-site by 20%, estimating a reduction in total freight costs by 8%

Oerlikon Fairfield Manufacturing, Lafayette, IN

Fall 2014

Continuous Improvement Intern

- Conducted work sampling of machines thereby increasing plant productivity and ensuring continuous improvement

- Formulated Master Production Schedule (MPS) template as per the requirements of raw and finished goods providing accurate inventory on-hand to obtain details of unreported scrap

- Programmed in Python to generate pie charts for each machine depicting the top 10% of the most efficient machines and the bottom 10% of the least efficient machines

Donaldson Company Inc., Frankfort, IN

Summer 2014

Manufacturing Engineering Intern

- Implemented Process Failure Mode and Effects Analysis (PFMEA) improving the quality and safety of the process

- Executed Cause and Effect analysis to investigate areas of improvement in all the assembly lines and providing solutions to line inefficiencies and scrap reduction projects

- Provided design structure and modeling for various safety issues and concerns in each line successfully using suitable Computer Aided Design (AutoCAD) software

- Developed Statistical Process Control (SPC) reporting formats thereby continuously improving various processes through statistical analysis

Callidai Motor Works Ltd., Chennai, India

Spring 2013

Engineering Intern

- Designed and developed a portable hand-driven tricycle-cum-wheelchair for the physically challenged reducing the floor area occupied by the prototype by 44%
- Piloted a survey with several physically challenged people obtaining ergonomically apt data thereby simulating the results using ANSYS software under suitable environmental conditions
- Investigated several quality testing techniques as per Bureau of Indian Standards (BIS) on the prototype obtaining satisfactory results

SKILLS

Design and Programming – SAS, JMP, R, Arena, @Risk, NetLogo, C++, AutoCAD, Pro/E, MATLAB, and ANSYS

***Languages** – *Proficient in 7 languages* - English, Spanish, German, Hindi, Telugu, Tamil, Kannada, and Sanskrit**

DATA ANALYTICS PROJECTS

Purdue University, West Lafayette, IN

Spring 2016

Predictive Analysis of Forest Fires and the Causal Factors

- Utilized several Data Mining (DM) Techniques to obtain accurate results of predicting forest fires on a dataset.
- Conducted a t-test comparison among the models and concluded the best configuration as the Bayesian Additive Regression Trees (BART) model presenting the least Root Mean Square Error (RMSE)
- Model ascertained of predicting the burned area of small and large forest fires, thereby proving useful for improving firefighting resource management

Purdue University, West Lafayette, IN

Fall 2018 - Present

Predictive Analytics to assess food donation behavior for food rescue and delivery operations

- Implementation of Statistical and Machine Learning algorithms to identify key predictors of in-kind food donations
- Compare the food supply pattern for three fiscal years to infer how food donation has evolved or changed
- Suggesting directions or options in decision making for reinventing operational effectiveness and refine the in-kind donation strategies to meet organizational objectives

Purdue University, West Lafayette, IN

Fall 2018 - Present

Predictive Analysis of Food Accessibility and Food Pantry

- Identification of food assistance deserts in the given region of Cleveland, Ohio thereby suggesting directions to refine food supply methods to such regions

LEADERSHIP EXPERIENCE

TEDxPurdueU, West Lafayette, IN

Fall 2015 - Spring 2016

Web Developer, Member of the Executive Board

- Led the web committee of the organization and collaborated with all the members of the board, gaining insights on the improvement of the TEDx website portal
- Pooled resources with Marketing and Events committee to enhance marketing strategies needed for a successful event

Society of Mechanical Engineers (SME), Chennai, India

Fall 2012 - Spring 2013

Core committee member

- Organized social events in order to promote innovations and creations in various fields of Mechanical Engineering
- Managed spending and usage of funds obtained for successful completion of engineering fairs and symposiums
- Organized seminars and talks held by engineering professionals on research-oriented topics of Mechanical Engineering

FIELD-RELATED EXPERIENCE

Purdue University, West Lafayette, IN

Fall 2017 – Present

Graduate Assistant, Technical Assistance Program (TAP)

- Providing technical assistance to enhance the productivity of private and public manufacturing, healthcare, and service enterprises in Indiana
- Implementing Lean Manufacturing tools and Design using Six Sigma philosophy for improving the business of these Industries

Purdue University, West Lafayette, IN **Fall**

2015 - Summer 2017

Graduate Teaching Assistant, Department of Industrial Engineering

- Used interactive teaching techniques to create an engaging and positive atmosphere providing assistance when needed
- Instructed students on the ARENA software utilized in their course projects ensuring good rapport and encouragement
- Led discussion sessions among the students ensuring sufficient mentoring to aid in the understanding of concepts
- Received excellent student evaluation scores and reviews that surpassed the course benchmarks for teaching quality

Burton D. Morgan Center for Entrepreneurship, Purdue University, West Lafayette, IN

Summer 2015 - Fall 2015

Graduate Research Assistant

- Designed a web-scrapping tool to organize and facilitate future compilation of a list of company details needed for the ‘Interns for Indiana Program’
- Conducted benchmarking analysis to assess the growth of the Purdue Entrepreneurship Certification Program in comparison to similar programs offered in the country
- Provided a literature survey to bolster the fact that communication and problem solving between engineers and entrepreneurs vary