Dynamic Demand Forecasting and Optimization: Improving the Management of Crisis-Driven Food Supply Chains

Christos Galanis

SID:

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JANUARY 2025

THESSALONIKI – GREECE

Dynamic Demand Forecasting and Optimization: Improving the Management of Crisis-Driven Food Supply Chains

Christos Galanis

SID:

Supervisor:

Supervising Committee Members:

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JANUARY 2025

THESSALONIKI – GREECE

Abstract

This dissertation, completed as part of the MSc in Data Science at the International Hellenic University, bridges theoretical insights with practical applications, providing value to individuals and organizations. The researcher conducted a comprehensive Systematic Literature Review to identify machine learning methods for food demand forecasting and EOQ techniques tailored to crisis-driven food supply chains. Using the Kaggle dataset "Food Demand Forecasting," an XGBoost model was employed to predict the last ten weeks of meal orders. These predictions, along with other parameters, served as inputs for a Linear Programming (LP) model designed to optimize order quantities under constraints such as budget and warehouse capacity. To enhance understanding, a budget reduction sensitivity analysis and demand variation simulation were performed to evaluate the impact of changing parameters or data. Finally, a user-friendly web application was developed, integrating the forecasting algorithm and LP model, offering a dynamic and interactive interface suitable for both technical and non-technical users.

Keywords: Crisis-driven food supply chains, demand forecasting, Economic Order Quantity, XGBoost, web application, Linear Programming, sensitivity analysis, simulation

Christos Galanis

14/01/2025

Contents

ABSTRACT III

CONTENTS V

TABLE OF FIGURES VI

TABLE OF TABLES VII

1 INTRODUCTION 9

1.1 PURPOSE OF THE STUDY 19

1.2 METHODOLOGY 20

1.3 OUTLINE OF THE THESIS 23

2 SYSTEMATIC LITERATURE REVIEW 24

2.1 SIGNIFICANCE OF LITERATURE REVIEW 25

2.2 IMPORTANCE OF SYSTEMATIC LITERATURE REVIEW 26

2.3 JUSTIFICATION OF HYBRID SYSTEMATIC LITERATURE REVIEW 27

2.4 PROCESS OF SYSTEMATIC LITERATURE REVIEW 28

2.4.1 Planning the Review 29

2.4.2 Conducting the Review 32

2.4.3 Reporting the Review 37

3 COMPREHENSION OF ARTICLES AND DISCUSSION 39

3.1 INTRODUCTION TO FOOD DEMAND FORECASTING 40

3.1.1 Research Question 1 Addressed 47

3.2 INTRODUCTION TO EOQ OPTIMIZATION 49

3.2.1 Research Question 2 Answered 51

4 QUANTITATIVE METHODS 52

4.1 FOOD DEMAND FORECASTING APPROACH 52

4.1.1 Objective 52

4.1.2 Methodology 53

4.1.3 XGBoost – Most Suitable Algorithm 60

4.1.4 Introduction of the dataset *64*
4.1.5 Exploratory Data Analysis (EDA) *65*
4.1.6 1st attempt with all features *72*
4.1.7 2nd approach with additional features *74*
4.1.8 3rd attempt with lag features *76*
4.2 LINEAR OPTIMIZATION METHOD 78
4.2.1 Application of Linear Programming *78*
4.2.2 Cost Reduction Sensitivity Analysis *84*
4.2.3 Demand Variation Simulation *87*
4.3 INTRODUCTION TO WEB APPLICATION 89
5 CONCLUSIONS AND RECOMMENDATIONS 99
6 BIBLIOGRAPHY 101

Table of Figures

Figure 1: Breakdown of Chapter 2 25
Figure 2: Research Scope 30
Figure 3: Distribution of publication years 38
Figure 4: Geographic Distribution of Study Focus 39
Figure 5: Merged dataset information 66
Figure 6: Total aggregated number of orders by week 66
Figure 7: Aggregated number of order for meal 1885 and center 55 67
Figure 8: Histograms and Boxplots of numerical attributes 68
Figure 9: Pearson and Spearman correlation matrices 70
Figure 10: Scatterplot Matrix 71
Figure 11: Total number of orders for each cuisine type 72
Figure 12: Feature importance plot using all features 73
Figure 13: Feature importance plot using additional features 75
Figure 14: Feature importance using lag features 77
Figure 15: Order Quantities and Unmet Demand Over Time for model first run 84

Figure 16: Budget Reduction Sensitivity Analysis Plot 87
Figure 17: Demand Variation Simulation Plot.. 88
Figure 18: Upload Datasets in Streamlit App 90
Figure 19: EDA in Streamilt App... 91
Figure 20: XGBoost Results in Streamlit App.. 92
Figure 21: Optimization Model in Streamlit App.. 93
Figure 22: Input Costs in Streamlit App.. 94
Figure 23: Budgets, Warehouse Capacities and Initial Inventories Inputs in Streamlit95
Figure 24: Optimization Results over Time by Meal and Center.......................... 96
Figure 25: Sensitivity Analysis Results in Streamlit App 97
Figure 26: Simulation Analysis Results in Streamlit App...................................... 98

Table of Tables

Table 1: Articles obtained after searching in each digital platform....................... 33
Table 2: Inclusion Process for Scopus 33
Table 3: Inclusion Process for ProQuest... 34
Table 4: Inclusion Process for IEEE.. 34
Table 5: Screening process for Scopus.. 35
Table 6: Screening process for ProQuest 35
Table 7: Screening process for IEEE... 36
Table 8: Hyperparameter grid applied to all features (XGBoost Documentation [Python API Reference], n.d.)... 56
Table 9: train file description 64
Table 10: fulfilment_center_info file description 65
Table 11: meal_info description.. 65
Table 12: Evaluation Metrics Result with all features model................................. 73
Table 13: Evaluation metrics using additional features 75
Table 14: Evaluation metrics using lag features 76
Table 15: Performance Metrics for main model 84
Table 16: Performance Metrics for Budget Reduction Sensitivity Analysis 86

Abbreviations: LP (Linear Programming), EOQ (Economic Order Quantity), SLR (Systematic Literature Review), GPU (Graphics Processing Unit), RAM (Random Access Memory)

1 Introduction

Imagine you are ordering your favourite meal online. With just a few taps, the restaurant prepares your dish, and within minutes, it's delivered right to your door—hot, fresh, and exactly as expected. Now, picture this happening on a larger scale: hundreds, thousands, even millions of people ordering their meals, all relying on an intricate network of food producers, suppliers, and distributors. Everything seems seamless—until suddenly, it isn’t.

Now, imagine walking into a supermarket, where shelves are fully stocked with everything you need. You casually fill your cart with fresh produce, pantry staples, and household essentials, confident that these items will always be available whenever you return. For many, this abundance and convenience is something taken for granted. However, this is not the reality for everyone.

Now, imagine for a moment, a world where even in underdeveloped crisis-prone areas, people could eat without any concern. Imagine communities where food is always available, hunger is non-existent, and the fear of shortages is no longer a reality. A world where food supply chains run smoothly, even in the most vulnerable regions, ensuring that no child goes to bed hungry, and every family has access to nutritious food. However, such scenarios, although ideal, are far from the truth in many parts of the world.

Stop imagining and return to reality. In both highly developed countries and underdeveloped regions, the reality of food supply chains is fraught with challenges. Even though access to food is a fundamental need, the equal distribution of food across the planet is not happening and as the population size increases not all people will be assured to have access to health and nutritious food as underlined by [1] mentioned in the paper of [2].

In well-off societies, food shortages due to disruptions—whether from pandemics, geopolitical conflicts, or natural disasters—highlight the fragility of global supply systems. In underdeveloped regions, these challenges are magnified, where food insecurity is a persistent issue and disruptions to the supply chain can have life-threatening consequences. For these communities, food is not just about convenience: it is a matter of survival. Conflicts, climate change, economic instability, and other crises exacerbate hunger and malnutrition, leaving entire populations without reliable access to food.

Food Systems

According to [3] food systems involve all the resources and actions necessary for producing, processing, marketing, consuming, and disposing of products from agriculture and forestry, along with the required inputs and outputs at each step [4].

In the paper of [4] are described the three fundamental elements of food systems which are food supply chains, environments, and consumer behaviour. These elements can be affected by environmental drivers of change, for example technology, innovation, and infrastructure: pollical and economic factors: sociocultural factors: and demographic factors [4].

As described by [4], drawing on the work of [5], supply chains play vital role inside food systems where they represent the activities and agents that move products from producers to end consumers, including the management of waste disposal.

In the paper of [4] the main food supply chain processes are described: in the beginning, food production involves agriculture up to processing and packaging: subsequently, in distribution are including processes such as storage and transportation: then, trade comes which encompasses food economics, which is directly impacted by supply and demand. From the other perspective, consumption includes aspects for example food availability, utilization and preservation. Last but not least, in the disposal part are involved processes such as decomposition, nutritional loss, and recycling [6], [7], [8] as cited by [4].

Either private or public sectors can support the food supply chain activities resulting in a very complex and specific decision-making processes at each chain as mentioned by [9] [7] in the paper of [4].

Disruptions occur in supply chains when there is a significant breakdown in the movement of food products from production to the final stage of consumption [10].

The supply chain risks, uncertainty and disruption are increasing [9]. Crisis disruptions can have serious economic impact and can be either natural or man-made events such as strikes, recession, natural disasters, terrorist attacks, financial crisis, breakdown of essential services for example electricity, water and many more [4].

Disruptions often lead to destruction, damage, suffering among people, and even loss of life [11]. When this happens, an imbalance occurs between demand and supply, price instability, fluctuations in operational capabilities, heightened uncertainty, decreased availability of high-quality products, and instability in operational systems [4]

The disruptions will have negative financial impact and adverse operating results in transportation costs, inventory shortages and delays in orders as discussed by [12], [13] and mentioned in the research of [4].

These unexpected disruptions impact significantly the food production, processing, and distribution through the whole supply chain network. Moreover, they affect the access, availability, utilization and stability of food security of products [7], [14]. Between 2008 and 2018, the least developed countries lost around 108 billion US dollars because of the decrease in agricultural and livestock production due to extreme disasters [15].

Therefore, food systems should be ready to ensure reliable access to safe food for all people, considering every time the unexpected events that may occur [16].

COVID-19

Coronavirus affected many industries for example agricultural and food, petroleum and oil, manufacturing, education and many others according to [17], [18] as referred by [19].

As mentioned by [20] the COVID-19 pandemic, which began in late 2019, highlighted the vulnerabilities in food supply chains (FSCs) worldwide, revealing their susceptibility to widespread disruptions.

In the paper of [20] it is noted that consumers' habits changed due to governments' strict lockdowns implementation and social distancing implication. This phenomenon is described as panic buying [21], [22]. To this effect, in the research of [23] cited by [20] a reference for a study in China is being made where it was found that consumers were willing to pay a 60.5% premium.

Consumption and stocking up food goods have increased in periods of isolation and quarantine imposed by governments [24]. This can be seen from the survey conducted by Veeck & Xie, (2020) in China between 15-23 February 2020 where they concluded that 58.6% of consumers were stocking up and buying large quantities of food and beverages.

Panic buying affected companies' insights and impacted materials, production and material planning. Additionally, the labor-intensive aspect of the food industry made it difficult for food processing to ensure consistent staff attendance [20].

This extreme consumption of goods [25],[26] led to empty shelves and sequentially to shortages in important goods in remote areas as described from [20].

According to [27] as mentioned by [20] there are a variety of factors that influenced the speed of the food processing and the affected capacity such as self-isolation, staff sickness and restricted mobility.

As mentioned by [24] food production has been impacted from lockdowns and restrictions. The [28] highlighted that Covid-19 had negative impact also in the food security. Closed and partial-operating businesses and restricted areas resulted in disrupted production schedules and decreased in human workforce.

The food industry had to implement safety measures to prevent contamination from COVID-19. Since COVID-19 can be transmitted via droplet particles, it has the potential to come into contact with food, surfaces, and the surrounding environments within the food industry and its supply chains [29], [30]. An example is highlighted by [28] as mentioned in the research of [31] where in the global hunger crisis between 2007-2008, 33 out of 105 countries applied food policy restrictions. This led to disruptions in the economy and in the lives of people in many countries.

In some countries human labor is not locally but people go to work from nearby areas. In United Kingdom at the beginning of the COVID-19 period the planted crops required human labor around 90.000 or they would be lost. Due to travel restrictions, workforce in agricultural areas and farms declined [24]. This country gets its labor from Easter European countries where due to the strict restrictions and lockdowns it was hard to move freely [32], [33] cited by [34].

Lockdowns in large countries such as China is severe since it is considered as an international partner in the global supply chain as described by [35] cited by [19].

According to [36] mentioned by [28] the disruption in international trade halted the export of food products and the import of essential raw materials. A variety of restrictions were applied such as expensive inspections, onerous rules, and protectionist acts from competitors [37] mentioned in the paper of [28].

As described in the paper of [24] significant disruptions occurred in the transportation networks-air, maritime, rail and land services due to strict restrictions. Furthermore, they highlighted that these disruptions focused mainly on delays in the distribution of food between suppliers, retailers and consumers.

A great example is in China where the first outbreak of COVID-19 took place very restrictive measures occurred by the government. This resulted in delaying necessary production inputs, slowing down the outflow of agricultural products, shrinking labor and operational capacities [38].

Regulations by many governments were imposed to cope with the spread of the virus [19]. The result of this as stated by [19] lead in a significant decline in production rates and fall in household incomes. Additionally, in their paper they noted that the price of necessary food commodities was increased such as rice, wheat, corn, and many others. This had a large impact in some developing countries from Africa due to their fragile food security systems as written from [39] mentioned in the paper of (Mostafa, Hussein, Elsheeta, & Romagnoli, 2024).

In Uganda, 2019, about 15% of banana production was lost because of the poor handling during transportation. Inadequate communication between suppliers and stakeholders is another problem that resulted in the loss production of banana.

Prior to the COVID-19 pandemic, middle-class households typically sourced 50% of their food from food services and the other 50% from supermarkets. However, with the onset of the pandemic, nearly all food purchases by these households shifted to groceries and supermarkets as noted by [40].

Therefore, businesses had to react with the unexpected demand despite operational and labor capacity constraints as described by [41] mentioned in the research of [24].

Climate change

Climate change is one of the top environmental issues that the world face and due to its complexity, it has gained popularity according to [42]. According to [43] report since 1850, the average combined temperature of land and oceans has risen by 0.06°C per decade. However, since 1982, the rate of warming has accelerated significantly, increasing at an average of 0.20°C per decade.

Over the past century, global warming increases due to greenhouse gas emissions, fast population growth, and the combustion of fossil fuel. In the paper of [44] a reference is made to the Climate Change (IPCC) Sixth Assessment Report which identified the human activities as the main factors accounting for the increase on Earth's average surface temperature.

[42] referenced the work of [45] who highlighted the effects of climate change including ocean warning, rise in sea level, rising frequency of tropical storms and rainfall. The consequences of these effects are drought, floods, migration, crisis in agricultural activities and health issues as discussed in the paper of [46] cited by [42].

Furthermore, [47] mentioned the significance of climate change as a global problem which can trigger severe weather events such as fires, floods and many more exposing the food supply chains.

An example of a supply chain disruption event was the Hurricane Katrina that happened in 2005. Around 2000 people died and crisis occurred on aid relief, food supplies, and uncertainty of shipping procedures took place at the port of Louisiana according to [48].

With the rise of extreme weather phenomena and the disruption of food supply chains, sustainable crop production is jeopardized, food insecurity becomes unstable, and public health systems are undermined as highlighted by [49],[50], [51] cited by [2]. This results in creating vulnerable production regions exposed to external shocks and reducing their ability to support a dependable food system as highlighted by [52]. Moreover, [53] underlined that the vulnerability of economic systems, along with the reliance on imports and exports of essential goods, leads to malnutrition.

[44] pointed out the direct and indirect impacts of global climate change on agricultural production. The direct impact focuses on crop cycles, yield and quality and imposes major challenges to stability, sustainability and livelihoods of farmers. On the other hand, the indirect effects concern on the change of the ecosystem. Due to the increasing frequency and intensity of extreme weather phenomena there is a scarcity and decrease in soil fertility, water availability, and changes in pests and diseases. The results create a decline in agricultural productivity and also raise food security issues.

Disruption in the logistics sector is critical too with challenges created in transportation, storage and warehousing [54], [55]. These challenges will be increased impacting the timely and efficient delivery of agricultural products [54] as cited by [56]. Crops vary to respond and survive in the climate change. This variation has implications for input costs and subsequently in market prices leading to economy disruptions. [56] drawing conclusions from [54] described that in developing countries food security and nutritional adequacy are at risk especially in areas vulnerable to extreme weather situations.

Russia-Ukraine War

After Covid-19 outbreak Russia-Ukraine conflict began which created even more challenges for Food Supply Chains. According to [57] both Russia and Ukraine are considered strong agricultural countries for commodities that include wheat, maize, barley, sunflower oil, and sunflower cakes. This is because they possess both countries large agricultural land. Based on [58] Russia is the biggest exporter of fertilizers in the world.

In the paper of [57] a discussion is being made regarding the six important impacts of Ukraine on the food supply chain which are the following. Firstly, a challenge is created in the food production, processing, and storage. [59] this situation can be attributed to labor restrictions and a reduced workforce availability, driven by conflicts and safety concerns. Therefore, activities of cultivation of soil and harvest of crops was reduced. The second impact was on transport logistics where food transportation interrupted due to export restrictions and port closures. Subsequently, [59] highlighted that there was a limited access in the agricultural land and its resources as cited by [57]. The latter described that the damaged transport infrastructure resulted in reduced access to farms and factories. [57] made a reference to the research of [59] where they highlighted importance of Ukraine as a major producer of wheat, maize, and sunflower. Therefore, the food supply chains between the Ukraine and European countries were at risk. The fourth impact was on the consumers. [59] underlined that there will be consequences on global health and food security. Example of this is the emerging countries in Africa and Asia which rely on Ukrainian wheat exports [57]. Furthermore, based on the paper of [58] food prices will be increased too due to the lack of different food commodities and fertilizers. The fifth challenge occurs in food-dependent services where most of the international refugees have lower their productions because either they are unlearned or new to work [57]. Also, according to [59] Ukraine is a global producer of fertilizers. Finally, a very important aspect of food is affected and that is its quality. Significant part on these plays different diseases for example that come from animals [57]. Moreover, the reduced use of insecticides and pesticides lead to a fall in crop yield [59] as mentioned in the research of [57].

[59] highlighted a decline in supply on a variety of food commodities for example cooking oil and wheat due to the sanctions and banning of trading of Russian and Belarussian agricultural products via the USA, European Union and their allies. Due to Russia's significant role as a gas supplier to Europe, the imposed sanctions have caused energy prices to soar by 3-4 times, leading to a sharp increase in food prices as well as mentioned in an article by [60] cited by [59].

Food Waste

According to a study conducted for the international Congress [61] mentioned in the paper of [62] it was calculated that more than a billion tons of food is either lost or waste. In other words, one-third of all food produced. Therefore, the food loss and waste are significantly high. Moreover, they have major impact on environment, social and economic areas such as the creation of greenhouse gases, global food security issues, and the costs associated with that based on the [63], [64], [65]. As described by [66],[67] food loss occurs at the early stage of the agricultural production whereas food waste at the end of the food supply chain.

Grain Security

[68] highlighted the importance of grain security which is a fundamental element of the economic development of a country since it provides social stability and national security. As they described, China plays a vital role in the world grain market as being the largest producer and consumer of grain on the planet. In their paper, pointed out that China's grain security is currently facing challenges such as water and soil resource shortages, a fragile ecological environment, and infrastructure limitations. However, there are a variety of challenges that China’s grain security faces such as globalizations, urbanization, climate change, COVID-19, regional conflicts and the objectives of “carbon neutrality” [69], [70], [71] as mentioned in the research of [68].

Aftermath of a disaster

In the aftermath of a disaster, the scarcity or mismanagement of relief supplies can severely hinder emergency response efforts, leading to increased suffering and delays in aid distribution. Accurate and timely information about the availability and logistics of relief supplies is crucial for effective humanitarian operations, as it ensures that resources are allocated efficiently, minimizing waste and reducing post-disaster losses. Moreover, precise demand forecasting and inventory management can play a pivotal role in optimizing supply chains during crises, helping to better prepare for and respond to future emergencies [72].

Food Security

Food security, defined at the World Food Conference in 1974, emphasizes that all individuals should have access to sufficient and nutritious food to meet their dietary needs according to [73]. Despite advancements, hunger and malnutrition persist, with undernourishment rates rising from 7.5% in 2017 to 9.3% in 2021 [74] as cited by [75]. Food insecurity manifests variably across regions, with developed countries facing distribution issues [76] and developing nations experiencing severe undernourishment, particularly in Asia and Africa [74] as mentioned in the paper of [75]. Hunger impacted 19.7% of the African population and 8.5% of the Asian people in 2021 according to [74].

The vulnerability of food security to economic and environmental shocks, such as the COVID-19 pandemic and geopolitical conflicts, has led to an alarming increase in hunger. In the paper of [75] is highlighted that due to the Russian – Ukrainian conflict approximately 345 million people are currently food insecure or will be at risk on falling into these conditions.

Understanding food consumption and production dynamics and implementing effective forecasting and optimization techniques are critical for enhancing food security, especially in crisis-driven supply chains as mentioned by [77]. In Kazakhstan is highlighted by [78] that food security has been identified as a primary objective in agricultural and economic policies as reported in the paper of [75].

Whether in urban centers or rural villages, supply chain disruptions expose vulnerabilities that ripple through economies and societies. Recent global events have intensified these challenges, from pandemic-related delays to geopolitical tensions, creating food shortages, fluctuating prices, and unpredictable supply. These crises make it clear that new, more resilient systems are needed to meet the growing demand for food in an increasingly uncertain world.

Food demand forecasting

Demand forecasting plays a significant part in supporting strategic organizational planning and decision-making processes according to [79]. Forecasts should be accurate and reliable since they are a key component of business decisions. Based on [80], [81], [82] as cited by [79] demand forecasting provides important insights for various industries.

Accurately forecasting food demand is a crucial challenge for both businesses and the pursuit of sustainable development [83]. Inside European Union 88 million metric tons of food are wasted annually, representing 15-16% of the European Union's food value chain as stated in the report [84]. In the research of [79] is highlighted the importance of accurately predicting food demand to prevent food waste. Demand forecasting includes two basic approaches: qualitative and quantitative. As reported by [85] the qualitative approach focuses on analyzing past performance, expert opinions, and judgments, while the quantitative approach utilizes historical data to predict future outcomes through mathematical modelling. Quantitative methods represent analysis of data relationships and patterns based on historical data to predict future trends. These methods can be predictive machine learning algorithms [86] as mentioned by [79].

Companies are focusing to accurate demand forecasting to meet the varying customer needs and to stay competitive in this unpredictable world [87]. Unprecise demand predictions can result to too much holding inventory: leading to high holding costs, risk of wastage or not enough inventory risking to out-of-stocks cases which translates as pushing the company's customers to search for its competitors [87].

Inventory Management

Many companies try to meet customer needs in the current global era where there are several problems that can occur and disrupt the inventory distribution process. One issue concerning companies is having a lack of raw materials and products [88] which lead to disruption of in production flows, rises the number of unfulfilled demand and increases the costs of inventory shortages. On the contrary, when inventory is too excessive to meet consumer demand then causes over stock leading to financial disruptions for the company, as the circulation of funds is slowed due to unsold inventory [89]. Therefore, the inventory should meet the input and output demand at optimal costs to be considered optimal as reported by [89]. Inventory management process should be implemented to optimize the inventory costs while considering the condition of the production being in stock. As highlighted by [89] having a good inventory management control a company can handle unexpected problems to meet demand. The objective of inventory control is to optimize production processes and ensure the availability of goods to meet demand at any given time [89]. In that regard, Economic Order Quantity (EOQ) is an inventory management system which calculates the optimal order quantity considering the minimization of the handling of inventory cost and the ordering cost [90].

In this new reality, the ability to predict food demand and optimize inventory has never been more critical. Accurate demand forecasting helps ensure that the right amount of food reaches those in need, minimizing waste and preventing shortages.

1.1 Purpose of the study

Due to the rise of unexpected events such as pandemics, wars, and environmental crises, disruptions in the food supply chain have become increasingly common, causing significant harm to both businesses and the global population. These disruptions underscore the urgent need for food supply chain managers to accurately predict demand and manage inventory efficiently, all while utilizing interactive, data-driven tools that can aid real-time decision-making.

The purpose of this study was to bridge the theoretical knowledge with practical applications. Therefore, a machine learning forecasting algorithm was developed to predict food demand and an inventory optimization method to determine the optimal number of orders under constraints tailored for crisis-driven food supply chains. All of these, integrated into an interactive web application that allows users for real-time decisions and thereby enhancing resilience, minimizing food waste and ensuring a more efficient response during crisis events. In addition, to conduct a Systematic Literature Review (SLR), this research explored different machine learning prediction techniques to predict food demand and an inventory management practice the well-known Economic Order Quantity (EOQ) within crisis-driven food supply chains and identified existing gaps.

There were three objectives presented in this research and are illustrated below:

1. To implement a machine learning forecasting model for food demand prediction.
2. To apply an optimization technique while considering various constraints. In addition, both constraint sensitivity analysis and simulation to be applied, assessing their effectiveness in managing inventory.
3. To develop a web application facilitating users in dynamically monitoring forecasts and inventory management, enhancing real-time decision-making in crisis-driven food supply chains.

The following two research questions were formulated:

1. What forecasting algorithms can be used for predicting food demand under crisis – driven supply chains?
2. How does inventory optimization, using EOQ, address the unique challenges of crisis-driven food supply chains?

1.2 Methodology

Research Design

This study employed a mixed-methods research design, combining qualitative and quantitative approaches. The qualitative component involves the SLR approach to evaluate existing research on machine learning food demand forecasting and EOQ in disruption-driven food supply chains and identify gaps in the literature.

The quantitative aspect analyzes a dataset. Initially, a machine learning forecasting model applied to predict food demand, evaluated using different regression metrics. Following this, a linear optimization technique was employed to determine the optimal order quantities, considering various constraints including warehouse capacity and budget constraints. In addition, budget sensitivity analysis and demand variation simulation scenarios were conducted to assess their impact on inventory decisions. The previously referred quantitative processes were integrated into an interactive web app.

Qualitative Component: Systematic Literature Review (SLR)

The researcher initially used the following comprehensive search string aimed at addressing simultaneously both research questions. However, it returned mostly non-relevant articles.

- (("Food Supply Chain" OR "Food Supply Chains" OR "Perishable Goods Supply Chain" OR "Fresh Food Supply Chain" OR "Food Sustainability Supply Chain") AND ("Crisis" OR "Disruption" OR "Food Shortage" OR "Emergency" OR "Crisis-Driven")) AND (("Demand Forecast" OR "Demand Prediction" OR "Machine Learning" OR "Algorithm" OR "Time Series") OR ("Inventory Optimization" OR "Inventory Management" OR "Cost Analysis" OR "Management" OR "Economic Order Quantity" OR "Operation Research" OR "Inventory Operation Research" OR "Inventory Simulation")).

This revealed a significant gap in the literature regarding demand forecasting and EOQ optimization for food supply chains under crisis conditions. Consequently, the researcher refined the focus, addressing individually the research questions using following two specific search strings:

- ("food demand forecast*" OR "food demand predict*" OR "predict* food demand" OR "forecast* food demand" OR "predict* demand food" OR "forecast* demand food" OR "demand forecast* food" OR "demand predict* food") AND ("Machine Learning" OR "Deep Learning" OR "Time Series")
- ("EOQ" OR "Economic Order Quantity" OR "EOQ model*" OR "EOQ method*" OR "Economic Order Quantity Model*" OR "Economic Order Quantity Method*") AND ("Food Demand" OR "Food Inventory" OR "Perishable Goods" OR "Fresh Foods" OR "Food Supply Chain")

The literature search was conducted across three digital libraries: Scopus, ProQuest, and IEEE. Inclusion criteria included full-text, peer-reviewed articles published from 2020 to 2024, covering journal articles, conference papers, books, and literature reviews, with a focus on open-access materials in English. The remaining articles were screened by title, abstract, and content.

Quantitative Component

Dataset

The Food Demand Forecasting dataset was sourced from Kaggle. It can be found on Kaggle website when searched for “Food Demand Forecasting.

Description

Originally created for a competition, the dataset consists of a training folder and a test file. This dataset is used mainly for predicting meal order quantities for various fulfillment centers. It contains a training set from week 1 to week 145. The test set is used for predictions from week 146 to 155. However, since we don’t have the answers for the testing set, we will neglect that file. Therefore, the focus will be on the training folder, which includes the below files:

- Train.csv: Contains information on food orders for various meals across fulfillment centers, including variables such as order quantities and fulfillment center IDs.
- Fulfillment_center_info.csv: Provides details about fulfillment centers meaning variables such as center IDs and city codes.
- Meal_info.csv: Contains data about meals, including variables for example category and meal IDs.

Food Demand Forecasting

The XGBoost model under Python framework was applied for its speed, computational efficiency, and robust handling of skewed data, non-normal distributions, non-linearities, outliers, and feature scaling due to its tree-based structure. Popular in machine learning competitions like those on Kaggle, Also, XGBoost provides feature importance weights through its Python library, allowing users to identify key attributes impacting model performance. This functionality was integrated into the web application too.

A variety of evaluation criteria implemented such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and R^2.

Linear Optimization

A Python-supported linear optimization process was employed to model the problem, incorporating constraints such as capacity and budget limitations. The performance metrics evaluated included:

- Total costs (ordering, holding, and penalty costs).
- Total unmet demand and its frequency.
- Budget utilization by center.

Budget Reduction Sensitivity Analysis

Sensitivity analysis was conducted to evaluate the impact of reducing the budget constraint on the optimal solution. The budget was systematically reduced in increments of 10%, 20%, 30%, 40%, and 50%, allowing the assessment of "how" and "how much" these changes influenced the solution.

Demand Variation Simulation

To analyze how variations in demand affect the optimization outcomes, simulation techniques were applied. Demand quantities were distorted at levels of 10%, 20%, 30%, 40%, and 50%. Each scenario was run through 100 simulations to ensure robust evaluation of the model's performance under varying demand conditions.

Development of Web Application

A web app was developed using Streamlit, an open-source Python framework, to integrate various components including EDA, the XGBoost model, an optimization model, and tools for sensitivity analysis and simulation. The app provides an interactive interface, allowing users to customize settings and parameters seamlessly, enhancing usability and enabling dynamic exploration of results.

1.3 Outline of the thesis

Chapter 1: Introduction

This chapter provides a brief introduction to food supply chain systems, discussing disruption events and the necessity for accurate food demand forecasting and inventory management under disruption events. Also, it outlines the purpose of the study, the objectives of the research, the research questions, and an overview of the methodology.

Chapter 2: Systematic Literature Review (SLR)

This chapter thoroughly describes the SLR approach, detailing the search strategies, and search strings used, the data sources and inclusion criteria applied, the findings from the literature review (LR), and the identification of gaps in existing research. Furthermore, the challenges encountered were described.

Chapter 3: Findings of Literature Review

In this chapter the extraction of relevant information from the remaining articles found by the SLR approach took place. In other words, the researcher made a thorough discussion on the findings of the papers to address the two research questions.

Chapter 4: Quantitative Methods

This chapter focused on the quantitative methods employed in the study. It includes a description of the dataset, EDA, application of the XGBoost model to predict food demand, the employment of a linear optimization technique. Additionally, budget sensitivity analysis and demand simulation were conducted. Moreover, it covers the development of the interactive web application.

Chapter 5: Conclusions

The final chapter summarizes the key findings of the study, discusses their implications for food supply chain management and offers recommendations for future research. Also, it highlights the limitations of the study.

2 Systematic Literature Review

A key part of an academic research is the literature review. In this chapter, a thorough Systematic Literature Review (SLR) was described to address the two research questions. In other words, based on the [91] and [92] a step-by-step approach methodology was presented in order to conduct the SLR process. The entire process was carried out with rational and well-justified reasoning for each action, allowing the reader to gain a clearer understanding of the procedure. The figure below illustrates a breakdown of this chapter:

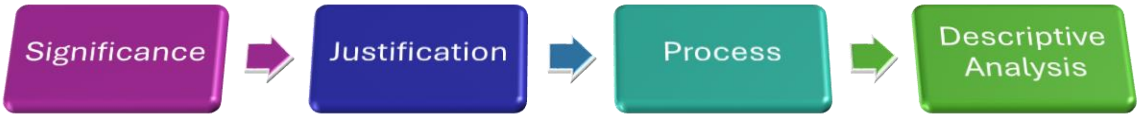

Figure 1: Breakdown of Chapter 2

As the figure above presents, the importance of the Literature Review (LR) process was being noted and the justification of why the researcher followed an SLR approach was presented. Moreover, the whole procedure was described thoroughly into steps. Finally, a descriptive analysis of the remaining articles took place in order the reader to gain a brief glance of descriptive data such as years of publications and geographical regions of the papers. In Chapter 3, the research questions answered after extracting relevant data from the remaining articles.

2.1 Significance of Literature Review

According to [93] many novice researchers mistakenly believe that simply summarizing papers is sufficient for conducting a literature review. Literature review is a necessary characteristic of academic research [91]. Hart (1998) as cited by [94] introduced a definition of the literature review as a bunch of ideas in the literature in order to justify a particular topic, the selection of methods and the presentation that the research results in something new.

A literature review goes beyond merely compiling papers or research manuscripts. Based on the paper of [91] to have knowledge improvement one must build on previously existing work. One should know where the current status is to push the knowledge frontier further. Literature review is a method that increases theory advancement [93]. By reviewing relevant literature, the researcher identifies the breadth and depth of the existing literature, detects gaps and unveils future potential research questions. In other words, the gap between past and present literature will be covered by future work by specifying hypothesis and / or develop new theories [91]. Also, one can assess the validity and quality of existing research using specific criteria, which helps identify weaknesses, inconsistencies, and contradictions within the body of work highlighted by [95] as cited by [91].

Literature Reviews should be valid, reliable and repeatable. Also, they should present thoroughly the methodology conducted. However, in many literature reviews we lack this characteristic in the planning field [91]. This technique is a non-systematic process of gathering, filtering and assessing the literature. Furthermore, it is biased on the writer's beliefs, knowledge and experience. Thus, the risk of prejudice is increased.

2.2 Importance of Systematic Literature Review

On the contrary, a systematic review points to gather evidence from research literature utilizing systematic and explicit processes, with the aim to answer research questions according to [96],[97] as cited by [98]. In that paper, is highlighted that a thoroughly methodology is followed to identify, gather, assess quality, analyze and discuss articles. However, since the articles will go through the screening process those remaining will be subjected to the researcher's actions. In other words, important literature may be missed since it is impossible to search in every database or due to the author's screening criteria.

[92] underlined the significance of Systematic Literature Review (SLR) method for its rigorous, structured approach, which enhances objectivity, reproducibility, and overall quality in reviewing literature. By following clearly defined, methodical steps, the SLR process minimizes bias, ensuring that findings are comprehensive and reliable.

In conclusion, the SLR method is a more systematic approach which follows steps for locating, gathering, assessing and writing from articles. This methodology is more suitable of answering scientific questions and achieving scientific objectives by a researcher as mentioned by [99].

2.3 Justification of Hybrid Systematic Literature Review

There are two forms of Literature Reviews [91] and these are described below:

- Those that support empirical research by providing contextual background: Background reviews are usually utilized to support research decisions, offer theoretical context or highlight gaps in existing literature that the study aims to address described by [100],[94] and cited by [91]
- Independent reviews: stand-alone reviews focus on synthesizing and interpreting existing research to provide a cohesive understanding of the topic reported by [101] and mentioned in the paper of [91]

In this thesis, the researcher focuses on stand-alone reviews which can be grouped into five categories as described below:

- Descriptive Review: This is the most common type of review, focused on examining the state of the literature surrounding a specific scientific question. It provides a summary without attempting to expand or build upon the existing body of work.
- Testing Review: Aimed at answering specific questions about the literature or testing a particular hypothesis, this review can take a quantitative or qualitative approach to analyze findings within the field.
- Extending Review: This review goes beyond summarizing existing studies; its purpose is to enhance the current literature by developing new, more advanced concepts or theories.
- Critiquing Review: In this type of review, studies are assessed against a predefined set of criteria. Rather than synthesizing the findings, this approach evaluates each study's alignment with established standards, identifying those that meet or fall short of these criteria.
- Hybrid Review: This combines elements of multiple review types, using an approach that integrates descriptive, testing, extending, or critiquing components as needed for a comprehensive examination of the literature.

In this study, a hybrid approach was employed that integrates both descriptive and extending review methods. This approach enables a comprehensive examination of the current literature to address the research questions (descriptive review). A significant gap emerged, as there was limited research focused specifically on "food demand forecasting" and "EOQ optimization for food" within crisis-driven food supply chains. This gap presented an opportunity to connect theoretical insights with practical applications in high-disruption settings. Therefore, the study expanded its focus to consider broader applications of food demand forecasting and EOQ, evaluating their relevance and potential adaptation for crisis contexts (extending review).

Both descriptive and extending methods have different variations. The author utilized the Narrative review and the Thematic synthesis respectively as reported below:

- Narrative Review is the most common type of descriptive review in planning according to [91]. Moreover, it is the least costly and rigorous regarding time and resources therefore it is a good option for selecting it [91]. However, it should be noted that this type of review is often biased to the writer’s experience, beliefs and overall subjectivity [91]. In this research, the author summarizes the key elements of each relevant article identified through the Systematic Literature Review (SLR) approach, focusing on those that align closely with the research question. This synthesis highlights critical insights and patterns from the literature, contributing to a comprehensive understanding of the topic.
- Thematic Synthesis is the process of extracting themes of the literature, clustering them and synthesizing them into analytical themes in order to address the research questions as mentioned by [102] in the paper of [91]. In this thesis, the researcher identifies similarities among the papers retained through the SLR method and synthesizes their content into a cohesive analysis. Key similarities observed include common datasets, applications, crisis contexts, and research scope.

2.4 Process of Systematic Literature Review

In the papers of [103],[104] as cited by [91] is highlighted that a review needs to include the following three stages in order to be considered successful: planning the review, conducting the review, and reporting the review. First, in the planning phase researchers recognize the need for a review, formulate the research questions and establish a review protocol. After that, the conduct of the review takes place in which the researcher identifies relevant primary studies, assess, extract and synthesize relevant data. Finally, the researcher writes the report to illustrate their findings from the literature review.

The steps undertaken in this SLR method draw on a combination of processes outlined in the works of [92],[91]. These steps are detailed below:

- Planning the Review: Define the research problem and the research questions, develop and validate the review protocol, including search strategies and inclusion criteria.
- Conducting the Review: Search the literature, screen studies, assess their quality, extract and synthesize data
- Reporting the Review: Provide a descriptive analysis of remained studies and report findings.

2.4.1 Planning the Review

Formulate the research scope

This review examined the critical role of food demand forecasting and inventory optimization (EOQ) within crisis-driven food supply chains, focusing on contexts impacted by wars, pandemics, natural disasters, and the challenge of minimizing food waste. An initial literature review revealed a significant gap in research directly addressing these methods in high-disruption environments, highlighting an opportunity to bridge theory with practical applications under such conditions. Therefore, this study expanded its scope to include general applications of food demand forecasting and EOQ in food supply chains, evaluating their relevance and adaptability in crisis scenarios. Emphasis placed on research utilizing machine learning, deep learning, and time series analysis to forecast demand, alongside EOQ-based models to optimize inventory. Studies exclusively related to non-food supply chains were excluded to maintain focus on food-specific challenges.

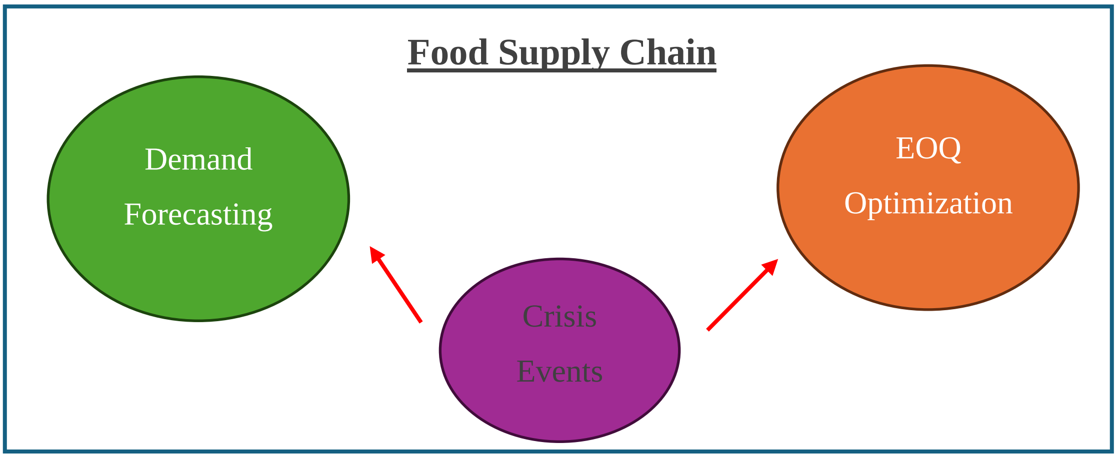

Figure 2: Research Scope

Develop and review protocol

Define Research Questions

There are two research questions that needed to be addressed with the SLR approach and are shown below:

1. What forecasting algorithms can be used for predicting food demand under crisis – driven supply chains?
2. How does inventory optimization, using EOQ, address the unique challenges of crisis-driven food supply chains?

Keyword Selection and Search String Development

Before the literature search in digital libraries began, it was essential to identify keywords that capture the information needed to address the research questions thoroughly. In this study, the focus areas were "food demand forecasting" methods, "EOQ optimization," and "crisis-driven food supply chains." Initially, the following search string was developed to encompass these three topics:

(("Food Supply Chain" OR "Food Supply Chains" OR "Perishable Goods Supply Chain" OR "Fresh Food Supply Chain" OR "Food Sustainability Supply Chain") AND ("Crisis" OR "Disruption" OR "Food Shortage" OR "Emergency" OR "Crisis-Driven")) AND (("Demand Forecast" OR "Demand Prediction" OR "Machine Learning" OR "Algorithm" OR "Time Series") OR ("Inventory Optimization" OR "Inventory Management" OR "Cost Analysis" OR "Management" OR "Economic Order Quantity" OR "Operation Research" OR "Inventory Operation Research" OR "Inventory Simulation"))

However, this initial search yielded many unrelated articles, highlighting a gap in the literature that specifically addresses food demand forecasting and EOQ optimization within crisis-driven food supply chains. This research gap underscored the need for targeted studies that focus on managing food supply chains during disruptive events using predictive food demand models and inventory EOQ optimization techniques.

To address this gap, the search strategy was refined to focus explicitly on "food demand forecasting" and "EOQ optimization" under the broader "food supply chain" theme, with an emphasis on crisis contexts. The revised search strings can be seen below:

- ("food demand forecast*" OR "food demand predict*" OR "predict* food demand" OR "forecast* food demand" OR "predict* demand food" OR "forecast* demand food" OR "demand forecast* food" OR "demand predict* food") AND ("Machine Learning" OR "Deep Learning" OR "Time Series")
- ("EOQ" OR "Economic Order Quantity" OR "EOQ model*" OR "EOQ method*" OR "Economic Order Quantity Model*" OR "Economic Order Quantity Method*") AND ("Food Demand" OR "Food Inventory" OR "Perishable Goods" OR "Fresh Foods" OR "Food Supply Chain")

Search Strategy Implementation

At this point, the research scope, research questions and the search strings had been constructed. The next phase included the selection of digital platforms used in academic research in order to search for studies and articles.

Data Sources selection

First, the Scopus database was used. Based on Wikipedia's information it is a scientific abstract and citation database developed by Elsevier. It indexes a wide range of peer-reviewed journals, conference papers, patents, and books across multiple disciplines. Also, it contains 82.4 million number of records [105]. Second, ProQuest is a multidisciplinary digital repository offering access to academic journals, dissertations, reports, and other research materials. Known for its extensive coverage of many diverse fields [106]. Finally, IEEE managed by the IEEE, IEEE Xplore is a digital library focused on engineering, technology, and computer science. It provides full-text access to IEEE's peer-reviewed journals, conference papers, and technical standards, making it essential for research in technical fields [107].

Inclusion Criteria Definition

The researcher used a variety of inclusion criteria. The aim of early screening was to filter out articles that do not align with the research questions or the established criteria [91]. At this point, it was important to be noted that each of the digital platforms allow different filtering criteria to be selected. After, careful consideration the researcher used the following criteria:

- Publications from 2020 to 2024 (Scopus, ProQuest and IEEE)
- Articles, literature reviews, books, and scholarly journals (Scopus, ProQuest)
- Full-text accessibility (ProQuest)
- Peer-reviewed publications (ProQuest)
- English language (Scopus, ProQuest)

Limiting the selection of publications to those from the past five years ensures that the thesis reflects the most current research and developments in the field. By focusing on recent studies, this approach provides relevant insights into modern challenges, technologies, and methodologies, while avoiding outdated perspectives. This enhances the relevance and accuracy of the analysis, grounding the thesis in up-to-date, credible knowledge.

Including articles, literature reviews, books, and scholarly journals in digital library searches provides a balanced foundation for a thesis. Articles offer recent research, literature reviews highlight trends and gaps, books cover foundational knowledge, and scholarly journals ensure credibility. This combination allows for a thorough, reliable, and well-rounded analysis. In contrast, sources like conference papers and theses are often preliminary or less rigorously reviewed, making them less reliable for foundational work. By prioritizing peer-reviewed and established publications, the thesis maintains higher academic rigor and credibility.

The remaining articles were screened by title and abstract. Then, those that met initial criteria were read in full so as the researcher to assess their quality and the most relevant ones addressing the research questions were selected.

2.4.2 Conducting the Review

In this part of the thesis, the processes of searching, filtering and assessing which articles will remain were described thoroughly.

Search the literature

Initially, using the two search strings defined earlier the author conducted two searches in each of the three digital platforms (Scopus, ProQuest and IEEE). The first search string focused on “food demand forecasting” algorithms while the second on “EOQ optimization” regarding food demand. In total, 487 articles remained as shown below:

Table 1: Articles obtained after searching in each digital platform

DATABASE	“Food Demand Forecast” Articles obtained	EOQ Optimization Articles obtained
Scopus	17	20
ProQuest	97	655
IEEE	15	5
Total	70	397

Inclusion Filtering

At this point, specific filtering criteria were applied to each of the digital platforms to weed out studies not related to the established inclusion criteria. The following tables show thoroughly the filtering process for each of the three digital platforms:

Table 2: Inclusion Process for Scopus

Inclusion Criteria For Scopus	“Food Demand Forecast” Articles remained	EOQ Optimization Articles remained
Years 2020 - 2024	15	8
Articles, literature reviews, books and scholarly journals	7	5
English language	6	5

Table 3: Inclusion Process for ProQuest

Inclusion Criteria For ProQuest	"Food Demand Forecast" Articles remained	EOQ Optimization Articles remained
Years 2020 - 2024	64	313
Full-Text / Peer - Reviewed	26	217
Articles, literature reviews, books and scholarly journals	26	216
English language	26	216

Table 4: Inclusion Process for IEEE

Inclusion Criteria For IEEE	"Food Demand Forecast" Articles remained	EOQ Optimization Articles remained
Years 2020 - 2024	13	3

From the tables above, it was concluded that the number of studies obtained from the inclusion filtering procedure for the "Food Demand Forecast" term were 45 whereas for the "EOQ Optimization" term were 224.

Evaluation of remained articles

At this stage, a screening assessment was conducted to ensure that the remaining articles align with the research questions. This process enhances the credibility and validity of the thesis by excluding irrelevant studies from consideration. To achieve this, a three-step evaluation process was applied, examining each article by title, abstract, and full text. The processes applied to each of the three digital platforms for the two search strings and illustrated in the tables below:

Table 5: Screening process for Scopus

Screening Processes For Scopus	"Food Demand Forecast" Articles remained	EOQ Optimization Articles remained
Title	6	Two (2) studies selected and the others were excluded such as "Energy implications of lot sizing decisions in refrigerated warehouses"
Abstract	6	2
Full-Text	Five (5) articles obtained leaving out the "Intelligent computational techniques of machine learning models for demand analysis and prediction" paper since it couldn't be accessed	2

Table 6: Screening process for ProQuest

Screening Processes For ProQuest	"Food Demand Forecast" Articles remained	EOQ Optimization Articles remained
Title	Twelve (12) remained excluding papers such as "A Geospatial Framework of Food Demand Mapping" and "Globalizing Food Items Based on Ingredient Consumption".	Eleven (11) papers obtained. Examples of articles excluded are the following: "Discontinuous Economic Growing Quantity Inventory Model" and "Inventory ordering policies for mixed sale of products under inspection policy, multiple prepayment, partial trade credit, payments linked to order quantity and full backordering"
Abstract	Five (5) articles found to be relevant excluding articles for example the "Application of neural networks in the	Two (2) papers remained. Instances of studies excluded are the following: Optimal economic order quantity model involving price discount" and "Retailer's optimal

	prediction of the circular economy level in agri-food chains" since it focused on predicting the circular economy and the "Machine learning application for sustainable agri-food supply chain performance: a review" where it described how machine learning helps in agri-food supply chain.	strategy for a perishable product with increasing demand under various payment schemes". Both of that examples didn't focus on food supply chains.
Full-Text	5	2

Table 7: Screening process for IEEE

Screening Processes For IEEE	**"Food Demand Forecast" Articles remained**	**EOQ Optimization Articles remained**
Title	Six (6) studies obtained and examples of papers rejected are "MangoTrace: An Intelligent Supply Chain Decision Support System" and "Voice Assistant for Fast Food restaurants using DistilBERT".	One (1). Articles excluded are the "Managing International Perishable Food Supply Chain: A Literature Review" and "Credit Policy Strategies for Green Product With Expiry Date Dependent Deterioration via Grey Wolf Optimizer"
Abstract	6	Zero (0). The only article left "Effect of COVID-19 on Supply Chain Decisions" was excluded since the EOQ models referred were not focused on food supply chains.
Full-Text	6	0

Data Extraction and Synthesis Process

The process of data extraction and synthesis in a narrative review is generally informal and lacks strict standardization. Typically, it involves collecting relevant information and presenting it narratively to provide both context and depth for the overall argument, which may carry an element of subjectivity [91]. In this thesis, each article was reviewed individually, the key elements of the papers were extracted and were summarized to align with the research questions.

For the thematic synthesis, themes of each paper were identified and organized into clusters (e.g. same dataset, application, scope) to address the research questions. Unlike traditional thematic synthesis methods that might use coding or tables [91], this approach did not involve formal coding or tabular formats. Instead, themes were grouped and refined in an informal manner to enable broader conclusions across the body of studies.

2.4.3 Reporting the Review

Report Findings

The reporting step emphasizes on transparent documentation of all processes of a literature review in order to be reliable and independently repeatable according to [108] mentioned by [91]. In that way, other researchers can follow the same steps and arrive to the same results. Inclusion and exclusion criteria [100], justification and rationale behind of each criterion [109], the key elements from the literature search, screening process and quality assessment [110] should be reported as well [91]. The process of the review should be made with a clear structure [111], transparent and conclusions need to be supported by data as mentioned by [91]. In this study, tables were provided to illustrate the inclusion filtering and screening process. By displaying graphically the conclusions, the comprehension of the review by the reader feels well-grounded highlighted by [112] in the paper of [91]. Moreover, every step of the SLR process that was followed was supported by rational justification.

Descriptive Analysis of the Articles

Overview

At this stage, a descriptive analysis of the remaining papers was conducted, grouping articles by publication year and geographic region. This approach enhances both the validity and interpretability of the findings by providing readers with a clear overview of the studies discussed in Chapter 3. Organizing the literature in this way helps contextualize research trends and regional focuses, highlighting potential gaps or patterns and offering a more nuanced perspective on the field.

Distribution of publication years

A total of 20 articles were identified through the SLR approach, all published within the specified five-year period outlined in the inclusion criteria. Notably, as shown in the figure below, 70% of these studies were published between 2022 and 2023, highlighting the recent and relevant nature of the research. While this ensures the review reflects the latest trends and advancements, it may also limit long-term perspectives and overlook foundational studies in the field.

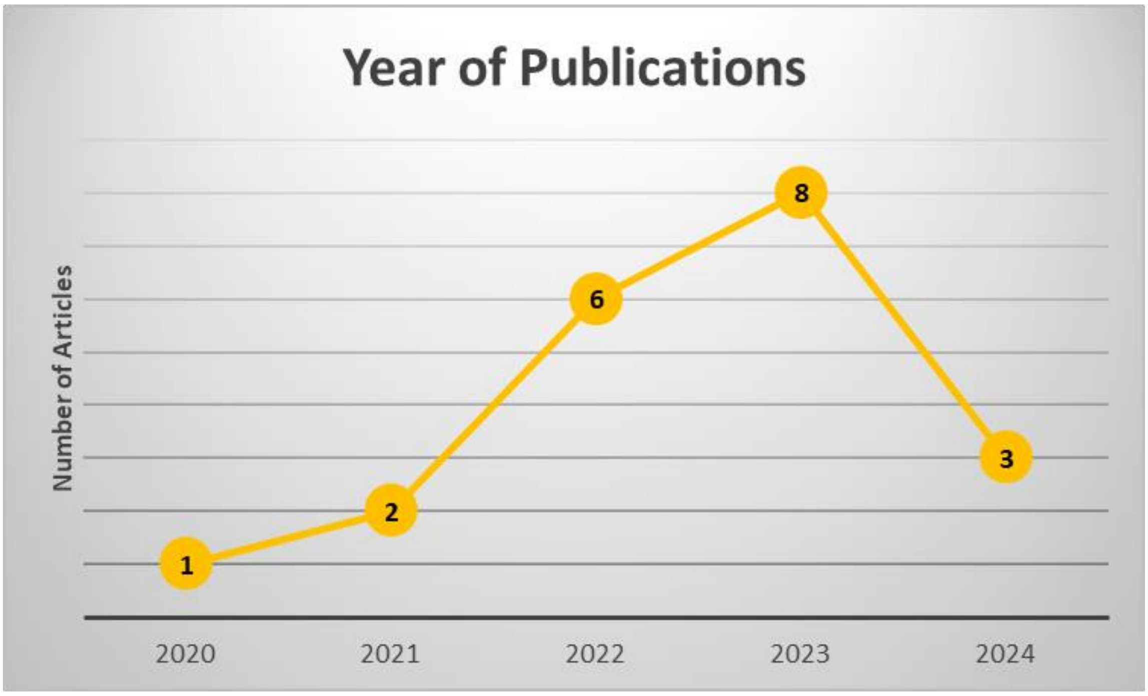

Figure 3: Distribution of publication years

Geographic Distribution of Study Focus

The distribution of countries in the studies is based on the dataset origin or the country where the application was implemented. 'Unspecified' refers to studies with datasets that lack a specified origin, while 'No Location' includes studies without a dataset or application tied to a specific country. Among the identified regions, China appears most frequently with 3 studies, followed by Portugal with 2. The remaining countries, such as

Turkey, Poland, and the USA, each appear once. This distribution highlights a strong focus on a few regions, while many studies lack clear geographic context, suggesting limited applicability to specific regions and potential gaps in geographic diversity.

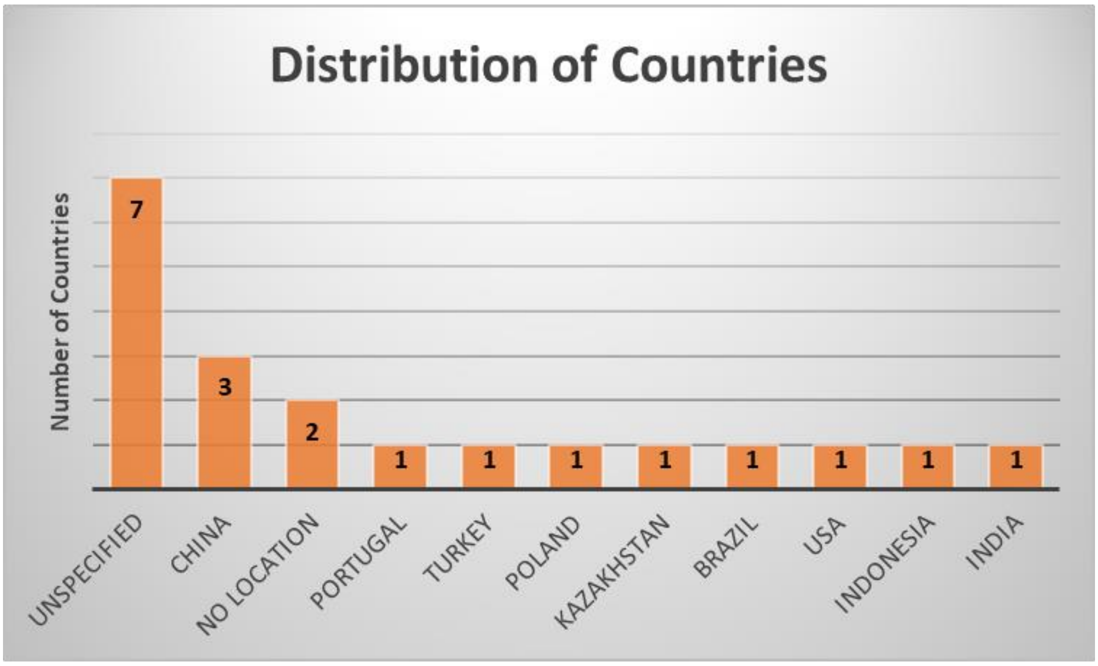

Figure 4: Geographic Distribution of Study Focus

3 Comprehension of Articles and Discussion

Overview

Despite the critical importance of food demand forecasting and optimization in crisis contexts such as wars, pandemics, natural disasters and geopolitical conflicts, a gap exists in the literature specifically focused on these scenarios. Few, if any, studies directly address demand forecasting or EOQ optimization tailored for crisis-driven food supply chains. To bridge this gap, this review includes general applications of food demand forecasting and EOQ optimization techniques, assessing their relevance and adaptability to crisis settings. By evaluating various case studies from the literature, the researcher highlights how these methods, though most of them designed for standard contexts, can be adapted to improve resilience and responsiveness in food supply chains during crises.

The author presented findings from articles identified through the SLR method, summarizing key insights from each paper in relation to the research questions. To enhance clarity, some articles are grouped based on similarities such as dataset usage and case study. This structure provides the reader with a cohesive, thematic overview of each paper, facilitating a clear understanding of the research questions addressed.

3.1 Introduction to Food Demand Forecasting

[113] reported that food companies face several challenges in accurately food demand prediction, largely due to the dynamic nature of consumer behaviour. Preferences and purchasing patterns are influenced by seasonality, weather, economic conditions, and emerging trends, all of which introduce uncertainty into demand predictions. This unpredictability can lead to inventory mismanagement, resulting in either overstock or shortages [113]. In crisis-driven supply chains—where disruptions can cause sudden changes in demand—such mismanagement is particularly detrimental. Overstocking not only leads to increased food waste and higher costs, but it can also exacerbate the challenges of maintaining profitability during periods of instability. Conversely, understocking can result in missed sales opportunities and dissatisfied customers, which is critical during crises when reliable access to food is essential. Developing accurate demand forecasting models is essential but complex, as it requires integrating historical demand data, seasonal patterns, and customer preferences into reliable predictive tools [113].

[113] highlighted several advantages of demand forecasting in the food industry, emphasizing benefits such as optimized inventory management, reduced food waste, and improved operational efficiency. Accurate demand predictions enable businesses to adjust staffing levels appropriately for peak and low-demand periods, which minimizes labor costs. Additionally, demand forecasting helps prevent overstocking and understocking, leading to better inventory control and a reduction in perishable waste. By ensuring the availability of popular menu items, demand forecasting also enhances customer satisfaction and loyalty, particularly important during crises when reliable access to food can impact community well-being. Overall, these insights support strategic decision-making, helping businesses adapt to consumer demand trends and environmental sustainability goals, especially in volatile contexts where flexibility and responsiveness are paramount.

Mainstream forecasting methods often fall short in the complex, fast-changing food industry [113]. Many businesses still rely on manual data entry and spreadsheets, which are time-consuming, prone to errors, and fail to capture all factors influencing demand. In crisis-driven supply chains, where rapid changes in consumer behaviour and external variables can occur, these outdated methods are particularly inadequate. Effective demand forecasting requires advanced systems that can handle unpredictable consumer behaviour, limited information, and external variables. By incorporating modern statistical models, machine learning, and real-time data, these systems can provide precise, timely forecasts that support optimized supply chains, inventory management, pricing, and promotions, ultimately enabling more informed decision-making in both normal and crisis situations [113].

Food Waste Scenario

Food waste is a major environmental and economic issue that comes mainly from poor demand forecasts according to [114]. In their paper, machine learning models are presented for short-term demand forecasting which has strong implications for crisis-driven food supply chains, where minimizing waste is essential due to limited resources and high demand volatility. In volatile environments like natural disasters, pandemics, or conflict zones, accurate forecasting can help prevent both overproduction (which strains supply during resource scarcity) and shortages (which can exacerbate food insecurity). By comparing random forest, LSTM, LightGBM, and transformer models, the study found that random forest and LSTM provided the best results across three different canteens. These models reduced food waste by up to 52% and improved forecast accuracy, which helps minimize both overproduction (a key contributor to food waste) and unmet demand.

Another study focused on minimizing food waste and improved resource planning was written from [79] which describes how to enhance food demand forecasting in university refectories using machine learning algorithms. The authors implemented various models, including Artificial Neural Networks (ANN), Gauss Process Regression (GPR), Support Vector Regression (SVR), Regression Tree (RT), Ensemble Decision Trees (EDT), and a Linear Regression (LR) on a dataset from Mersin University. The EDT Boosted model showed the best performance, significantly reducing prediction errors and thus supporting efficient production planning to address food waste issues.

A paper focused on developing a *Nonlinear Autoregressive Exogenous Neural Network (NARXNN)* model for food demand prediction to optimize inventory and reduce food waste was introduced by [83]. The study aimed to improve forecasting accuracy, specifically for processed foods like bread and butter. NARXNN was chosen for its ability to model time-series data by considering past values and external influences, which helps capture complex demand patterns. The model's high prediction accuracy can significantly enhance inventory management, enabling food suppliers to respond quickly to sudden changes in demand. This is especially important in high-disruption contexts where traditional forecasting methods may fall short. However, a notable limitation of the NARXNN model is its struggle with small datasets, particularly those with fewer than 100 rows, which could be a concern in certain crisis situations where data availability is limited. Nevertheless, when applicable, this model represents a valuable tool for promoting sustainable practices and improving the resilience of food supply chains in challenging environments.

These approaches showcase the potential of machine learning and deep learning to enhance sustainable food demand forecasting, addressing the critical issue of food waste and reducing environmental impact.

Food Security Scenario - Kazakhstan

The study by [75] developed a neural network model to forecast food consumption trends in Kazakhstan, specifically addressing food insecurity exacerbated by external factors such as migration from the Ukrainian-Russian war and inflation. By utilizing principal component analysis (PCA) to identify key socioeconomic indicators—like poverty rate, GDP per capita, and food price index—the researchers created two artificial neural network (ANN) models that achieved high accuracy and low prediction errors.

In crisis scenarios, such as the one currently affecting Kazakhstan, accurate forecasting of food consumption trends is essential for addressing food insecurity. The model's projections indicate a decrease in total food consumption over the next three years, along with significant shifts in consumption patterns, including decreases in potato and meat consumption and increases in bread and dairy products [75]. These insights are critical for policymakers and food supply chain managers to adjust their strategies accordingly, ensuring that food availability aligns with evolving consumer needs and mitigating potential shortages. By enhancing the ability to predict changes in consumption due to crises, this approach can support more effective resource allocation and planning, ultimately fostering greater food security in affected regions.

Relief Supplies Scenario – São Paulo, Brazil

The study of [72] underscores the importance of accurately predicting the demand for relief materials to inform decision-making in advance of natural disasters. Utilizing a dataset of aggregated demand for relief supplies from the municipalities of São Paulo, Brazil, between 2015 and 2020, the researchers found that around 80% of this demand focused on food baskets and cleaning kits. The researchers applied three predictive models: Negative Binomial (NB), Zero-Inflated Negative Binomial (ZINB), and Zero-Inflated Negative Binomial Multilevel (ZINBM). The ZINBM model provided the best fit, achieving the lowest Akaike Information Criterion (AIC) and log-likelihood values, as well as yielding the most accurate estimates among all models tested.

In crisis scenarios, such as natural disasters, the ability to predict the demand for essential supplies is critical for effective resource allocation and timely delivery of aid. The ZINBM model's success illustrates the advantages of Generalized Linear Mixed Models (GLMM) in handling the complexities of nested data, which is often present in such contexts. By providing proactive assessments of relief supply needs, this model can enhance preparedness and response efforts, ensuring that food and essential supplies are available when communities need them most. This proactive approach is vital for mitigating the impact of disasters, improving overall resilience in affected regions, and ultimately helping to secure food access and safety during emergencies.

Grain Consumption Scenario – China

The studies by Zhu, Wang, and Tian (2020) and Zhang et al. (2022) focused on forecasting grain consumption, each offering valuable insights for crisis-driven food supply chains.

[115] introduced an enhanced multivariate linear regression (MLR) model integrated with time series forecasting. This study improved accuracy by refining the selection of impactful factors, using historical data as a primary factor. The methodology incorporated grey relational analysis and Pearson correlation to select key predictors, optimized data preprocessing with moving average filtering to stabilize fluctuations, and combined MLR

with time series data to capture internal trends. The paper also compared this modified model's performance to traditional MLR, ARIMA, and machine learning models like BP and LSTM. Making full use of the advantages of the time series and the MLR model results showed that the proposed model significantly improved accuracy, achieving low error rates across various timeframes, especially in short- and medium-term forecasts, making it highly applicable for forecasting in agricultural sectors where accurate predictions can inform resource allocation and planning during crises, helping to mitigate supply chain disruptions.

In contrast [68] used the Grey Model (1,1) (GM) to accurately forecast China's grain consumption from 2022 to 2031, highlighting its effectiveness for systems with limited data and uncertainty. The model predicted a steady 1.53% annual increase in grain demand. The findings suggested that the GM(1,1) model is an invaluable tool supporting policies aimed at resource efficiency and food security, essential for preparing for future challenges and ensuring stable food supplies during crisis situations. Together, these studies illustrate the importance of accurate grain consumption forecasting in enhancing food security and resource management in times of disruption.

Food Banks - Scenario

The paper of [116] proposed an advanced demand forecasting model for food banks such as local supermarkets, manufacturers and community organizations to predict donation volumes more accurately. By clustering donors according to behaviour-based attributes like reliability, affiliation, donation quality and more, the models used K-Means and K-Medoids segmented donors into profiles, which then fed into an ensemble forecasting approach. The authors used individually the six following models: Moving Average (MA), Naïve Forecasting Method (NFM), Seasonal Naïve (SN), Exponential Smoothing (EM), Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR). Moreover, they applied an ensemble method which averages the six previously referred model's forecasts and in that way the accuracy of predictions was enhanced significantly compared to standalone methods. Results indicated that the ensemble method boosts forecasting precision. In crisis-driven contexts, where demand for food banks may surge unpredictably, this model's ability to improve forecasting precision is highly valuable. Accurate demand forecasts enable food banks to optimize logistics and allocate resources more effectively, ultimately enhancing their capacity to meet demand and reduce food insecurity during emergencies.

Forecast of fresh food demand scenario – Turkish supplier

The paper of [117] described a machine learning-based approach for daily demand forecasting in the fresh food retail market, essential for managing inventory, minimizing waste, ensuring food availability and delivering fresh food to customers. Using a dataset from a large Turkish fresh food supplier, the researchers evaluated models such as Long Short-Term Memory (LSTM), Transformers, Prophet, Feedforward Neural Networks (FNN), and eXtreme Gradient Boosting (XGBoost). Each model leverages unique strengths, with LSTM adeptness at processing sequential data and its capability to capture temporal patterns and long-term dependencies. Transformers handling longer dependencies, and XGBoost effectively capturing nonlinear relationships and its robustness to overfitting. Experimental results showed that while all models perform well, XGBoost slightly outperformed others, offering stability and lower variance in predictions. The study highlights that accurate forecasting can enhance supply chain efficiency.

Food Meal Delivery Companies – Case study

The studies by Panda & Mohanty (2023), Pandey et al. (2023), Anitha et al. (2023) and Chatterjee (2022) demonstrated the effectiveness of machine learning and deep learning models in forecasting food orders within supply chains, especially in meal delivery services managing diverse products across multiple distribution centers.

[87] demonstrated the effectiveness of machine learning and deep learning models, including Random Forest, various Gradient Boosting techniques, and LSTM models. Their approach included feature engineering techniques such as lag features and Exponentially Weighted Moving Averages (EWMA) to improve model accuracy. Results indicated that LSTM models are especially effective in capturing long-term dependencies, enhancing demand forecasting accuracy during periods of volatility, reducing food waste and ensuring better resource allocation. Similarly, [118] applied Random Forest and Gradient Boosting Regressors, along with preprocessing steps like feature encoding, log scaling, lag feature creation, and rolling mean calculations. Their findings showed that the Gradient Boosting Regressor yielded the highest predictive accuracy aids in stock planning,

crucial for adapting to sudden shifts in demand during crises. [113] utilized the Random Forest algorithm which aggregates predictions from multiple decision trees, increasing accuracy by reducing overfitting and enhancing generalization. Results indicated precise forecasting ability. To facilitate practical application, a Flask web application was developed to provide easy access to demand forecasts and offer a user-friendly interface for visualizing model performance. In crisis-driven supply chains, this approach could be highly beneficial. The accessible web application allows supply managers to quickly check everything dynamically, reducing waste and ensuring consistent food supply—key for resilience and effective response during crises. Another study introduced a stack-based model with an explainability focus for food demand forecasting [119] enhancing decision-making in crisis situations. Traditional machine learning models like linear regression, decision trees, and XGBoost were combined into a stacked ensemble model to improve predictive accuracy, with a meta-model integrating these base predictions. To address the interpretability of this complex model, Local Interpretable Model-agnostic Explanations (LIME) was used to highlight feature importance, showing that factors such as base price, checkout price, and promotional elements significantly impact demand predictions. The LIME analysis provided insights into how each feature influences outcomes, enhancing the model's transparency and allowing for informed decision-making in food supply management in high-pressure contexts [119].

Prediction of high value-added agri-foods – Case study

[120] presented a hybrid demand forecasting method for high value-added agri-foods, integrating machine learning and time series analysis that can be useful for crisis-driven food supply chains. The authors created a fusion model combining machine learning models which can handle high-dimensional data with the ARIMA algorithm which captures trends and cyclical information to enhance predictive accuracy. Using annual data from Zhejiang Province (2006-2022) with many dimensions and limited data, the authors selected key economic indicators, such as population and GDP, as predictors. The model followed a Stacking approach, using machine learning algorithms, ARIMA performed rolling forecasts to adapt to recent trends, and Ordinary Least Squares (OLS) regression optimized the combined predictions. This fusion model reduced errors by 25% compared to standalone models, demonstrating improved precision in scenarios with limited data. This study underscored the effectiveness of hybrid approaches in forecasting demand for

high-value, perishable agricultural goods, contributing to resource-efficient supply chain planning [120]. This hybrid model can be utilized by supply chain managers in periods of unexpected events where disruptions can lead to volatile demand and shifts in order to adapt efficiently to the new changes even with limited data.

A heterogeneous model approach

The method by [121] presented a meta-framework for enhancing forecasts of heterogeneous time series for instance different seasonality or trends, applied specifically to food demand forecasting. It is highly applicable to crisis-driven food supply chains due to its adaptability to irregular, short, or sparse time series data, which are common in volatile environments like wars or natural disasters. By using dynamic time warping (DTW) to identify similar time series and k-nearest neighbors (k-NN) to combine forecasts, it improved accuracy even with limited data. Moreover, it is adaptable to both simple and complex models and leverages similarities across diverse time series to improve forecast precision. This flexibility allows it to adapt to both regional and product-specific variations in demand, making it well-suited for responding to sudden shifts and high disruption, ultimately enhancing resource allocation and minimizing waste in crisis scenarios.

Limitations of deep learning

While deep learning techniques have shown promise in food demand prediction and inventory management, several challenges persist [122]. First, these models depend heavily on high-quality, available and large-scale labeled data for training. Additionally, handling sensitive customer data raises concerns about privacy and security. Another major challenge is the interpretability and explainability; their complex architectures make it difficult to understand and explain the decision-making processes involved. Finally, the computational demands of deep learning models can limit scalability and hinder real-time deployment, especially for large-scale food delivery systems.

3.1.1 Research Question 1 Addressed

- What forecasting algorithms can be used for predicting food demand under crisis – driven supply chains?

Food demand forecasting in crisis-driven supply chains relies on both traditional and advanced models. Traditional methods, including multiple regression, exponential smoothing, Holt-Winters, ARIMA, random forest, and gradient boosting, are widely used to detect patterns in historical data. However, these models are limited in rapidly changing markets as they require frequent redevelopment and lack adaptability to new data [83]. Random forest, for instance, achieves accuracy by aggregating decision trees, reducing overfitting and enhancing generalization [113].

To overcome these limitations, machine learning and deep learning models have been introduced for greater flexibility. RNNs (particularly LSTM and GRU) capture complex temporal demand patterns and non-linear relationships, while CNNs analyze spatial patterns, considering factors like population density and cultural influences. Additionally, GNNS assist in optimizing supply chain networks while attention mechanisms help models account for contextual factors [122]. Transformers are effective for long-term dependencies, while XGBoost captures nonlinear relationships with robustness to overfitting [117].

Specialized models also address specific needs: NARXNN captures past values and external influences but struggles with limited data [83], while Generalized Linear Mixed Models (GLMM) manage nested data, common in crisis contexts [72]. Hybrid models, like multivariate linear regression (MLR) integrated with time series, enhance predictive accuracy (Zhu et al., 2020), and ensemble learning, combining models like Moving Average (MA), Naïve Forecasting Method (NFM), Seasonal Naïve (SN), Exponential Smoothing (EM), Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR), boosts precision by averaging predictions [116]. Stack-based models like linear regression, decision tress and XGBoost and fusion approaches such as machine learning with ARIMA integrate diverse model outputs for more robust forecasts [119],[120]. Dynamic time warping (DTW) and k-nearest neighbors (k-NN) further improve accuracy by leveraging time series similarities using both simple and complex base models [121]. Finally, interpretability tools like LIME offer transparency in complex model predictions, crucial for decision-making in crises [119].

3.2 Introduction to EOQ optimization

Inventory management is a crucial process in the food industry, helping minimize food waste across the supply chain, from warehouse receipt to final customer delivery [123]. With a rising population, the demand for efficient inventory management has increased to ensure a steady supply of stocked food items. Many food companies experience losses due to food waste at various stages, including spoilage, transportation issues, and weather conditions [123]. To operate effectively, companies must maintain adequate raw material inventories to meet production demands [124]. [125] note that inventory investments can range from 20% to 60%, underscoring the importance of balancing inventory levels.

Effective stock control encompasses purchasing, receiving, warehousing, and restocking activities, all of which help reduce operational risks [124]. [126], as cited in [125], reported that excessive inventory levels can drive up storage and maintenance costs in warehouses. Conversely, inefficient inventory management—especially in cases where a company runs out of stock—can lead to production disruptions and hinder smooth operations.

Effective inventory management is essential for meeting company demands while minimizing potential risks. A widely used technique in inventory management is the Economic Order Quantity (EOQ) model, which helps companies determine the optimal order quantity to meet their needs efficiently [125]. Furthermore, to optimize inventory, demand forecasting is used, though it is only an estimate and may not always match actual needs. As a result, companies employ a safety stock strategy to maintain competitive performance. However, maintaining safety stock incurs additional capital costs, making it essential to strike a balance between service levels and capital costs [125].

Small-Scale Indonesian food enterprise scenario

The study of [124] explored inventory control optimization for perishable goods, specifically soybeans used in tofu and tempeh production at UD XYZ, a small-scale food enterprise. Faced with capital constraints and limited warehouse capacity, traditional EOQ methods were inadequate, leading to excess costs and storage issues. The study proposed a modified Economic Order Quantity (EOQ) using the Lagrange multiplier to account for these limitations. Results indicated that while the standard EOQ suggested lower total inventory cost, the Lagrange EOQ balanced constraints by reducing the order size to meet the warehouse capacity, maintaining an optimal balance with inventory costs. This modified EOQ approach is particularly valuable in crisis-driven supply chains, where disruptions often lead to limited resources and space. The model's adaptability to spatial and financial constraints makes it a viable solution for optimizing perishable inventory in restricted environments, ensuring efficiency and reducing waste during periods of high uncertainty.

Closed-form EOQ solutions – Introduction

[127] highlighted approaches on managing inventory items that degrade over time, such as food, using the Economic Order Quantity (EOQ) model adapted for exponential deterioration. This holds significant relevance for crisis-driven food supply chains where logistical challenges and supply chain disruptions can lead to extended inventory holding times and therefore the risk of inventory deterioration increases. [127] described that traditional EOQ methods, which assume items remain viable indefinitely without any physical losses, are inadequate for items that deteriorate at a rate proportional to their current inventory level. Existing approximations of the EOQ model yield closed-form solutions that are not always accurate or intuitive. This study introduced six different closed-form solutions for both the basic EOQ model and its backordering variant for deteriorating items, demonstrating improved accuracy in terms of optimal order quantity and the optimal total cost and ease of use for practical applications. The solutions balance ordering, holding, and deterioration costs effectively, contributing to sustainability by reducing waste in supply chains. The paper provided extensive experimental results, showing the proposed solutions' superiority over existing methods and complete proofs to show that the proposed closed-form solutions are unique optimal solutions, making them highly applicable for managers and decision-makers in industries facing perishable inventory challenges.

Dried food hospital case study

The paper of [125] provide valuable insights into inventory cost reduction for crisis-driven food supply chains, particularly in high-stakes environments like hospitals, where food accessibility is essential, and resources may be limited. The study revealed that existing practices led to excess stock levels, averaging 39.3% above necessary quantities, resulting in higher storage costs and a 29.73% loss due to spoilage. Using EOQ, the authors [125] calculated optimal order quantities, safety stock, and reorder points to improve efficiency. The results showed that EOQ application reduced total inventory costs by 44.22%, demonstrating its effectiveness in minimizing expenses and addressing overstock issues in the hospital's food supply which are critical in crisis scenarios. In high-disruption conditions, where accurate demand forecasting and optimized ordering are crucial to meeting fluctuating demands and avoiding waste, the application of EOQ could improve responsiveness and resilience, ultimately ensuring more consistent food supplies.

Fish products EOQ case study under two scenarios

This study of [123] addressed optimal inventory management for perishable fish products under two scenarios: normal backordering and lockdown-induced backordering, where demand and shortage costs differ significantly. It developed an inventory model that minimizes total costs, considering holding, shortage, and deterioration costs, with a penalty cost introduced during crisis conditions meaning lockdown to obtain the optimal total cost. Sensitivity analysis showed that as shortage parameters, penalty costs, and demand variance change, the inventory levels and total costs fluctuate. The model effectively reduced costs by adjusting backordered quantities and supporting efficient supply chain management from wholesalers to retailers during volatile demand periods.

3.2.1 Research Question 2 Answered

- How does inventory optimization, using EOQ, address the unique challenges of crisis-driven food supply chains?

The studies reviewed explored the optimization of inventory management for perishable goods within food supply chains using tailored Economic Order Quantity (EOQ) models while the author emphasized the importance of those models on how to meet crisis-driven demands. [124] adapted EOQ for small-scale food enterprise with capital and space constraints, enhancing inventory balance in case of disruptions. (Çalı¸skan, 2022) introduced closed-form EOQ solutions for deteriorating items, crucial for crises where holding times are prolonged, ensuring lower waste and costs. [125] demonstrated EOQ's cost-saving potential in a hospital setting by minimizing excess inventory and spoilage, vital during emergencies. [123] applied EOQ to fish products under normal and lockdown conditions, adjusting for demand and shortage cost variances to reduce overall supply chain costs. Together, these EOQ adaptations support resilience and efficiency in food supply chains amid volatile demand and resource constraints.

4 Quantitative Methods

This chapter includes the following elements:

- Food Demand Forecasting Approach
- Linear Optimization Method
- Web Application Development.

4.1 Food Demand Forecasting Approach

4.1.1 Objective

The primary objective was to forecast the number of meal orders across various centers for the last 10 weeks of the available dataset (i.e., weeks 136–145).

Note:

To achieve the objective, various methods were explored, drawing insights from scientific papers (via a systematic literature review) and Kaggle notebooks. However, only XGBoost model was selected and applied to the dataset for a variety of reasons that were explained later. Key findings were documented and clearly presented for reader comprehension. Since the original competition targeted demand prediction for weeks 146–155, but data for those weeks is unavailable, the focus was shifted to predicting demand for weeks 136–145.

4.1.2 Methodology

This dataset can be approached as either a standard machine learning problem or a time series problem. Although it includes a time-related variable (e.g., weeks), this variable alone is insufficient to extract more complex time-dependent features from the existing data. However, the presence of the "week" feature cannot be ignored entirely.

The dataset involves a variety of meals distributed across different centers over multiple weeks. Unlike traditional time series problems, where a single time-dependent entity (e.g., one meal in one center over time) is analyzed, this dataset encompasses multiple entities simultaneously. As a result, it does not fit neatly into the framework of a standard time series problem.

To address these challenges, a hybrid approach was adopted. The problem was treated as a machine learning task, focusing on the dataset as a whole rather than isolating individual time series, such as a specific meal in a specific center. However, the time-ordering aspect was not discarded. The idea of temporal features like lagged values was incorporated to retain the dataset's temporal structure. This approach allowed for a comprehensive integration of machine learning and time series methodologies, enhancing the ability to capture the dataset's complexity effectively.

The analysis was conducted using the Python programming language within Google Colab, utilizing a paid version with enhanced GPU and RAM resources.

The notebook was structured as follows to effectively address the objective:

The process began with a thorough EDA to gain a clear understanding of the dataset and then model development took place where four approaches were implemented to progressively enhance the model's performance. These are detailed below:

1. Integration of all available datasets into a single dataset.
2. Addition of new features like dataset-wide features.
3. Introduction of features such as lag variables.

4. The most valuable features, most suitable techniques, and optimal hyperparameters were chosen for building the final model for integration into the web application.

Throughout these approaches, pre-processing, feature selection and model evaluation techniques were constantly applied to refine and assess the model's performance. Specifically:

- Preprocessing: Involves transforming raw data attributes into a format understandable by computers and machine learning algorithms [128]. For categorical data, this process is known as encoding.
    - Label Encoding assigns a unique integer to each category [87], suitable for ordinal data [129].
    - Binary Encoding, which uses bitstrings to represent categories [130], is effective for high-cardinality features as it reduces dimensionality, unlike One-Hot Encoding. However, it can introduce order among categories and may lead to information loss, potentially reducing model performance [131].
    - One-Hot Encoding, which creates a new feature for each unique category each binary [87], [132] increases dimensionality capturing all the information and removes any implied order.

  [87],[118] applied Label Encoding in the same dataset employed in this thesis. However, as the categorical features lack any inherent order, this thesis instead applies One-Hot Encoding, using the OneHotEncoder transformer provided by the scikit-learn library [133], to transform these features into numerical representations.

  To prevent data leakage, the OneHotEncoder was fitted on the training set and then applied to the test set, ensuring that unseen categories in the test data did not influence the training process. A simple example to illustrate data leakage in the context of One-Hot Encoding is the following:

  Suppose the training dataset contains a categorical feature *City* with values *New York*, *Los Angeles*, and *Chicago*. The test dataset includes an additional unseen category, *San Francisco*. If the One-Hot Encoder is fitted on the combined training and test datasets, the *San Francisco* category would influence the training process, even though it is not present in the training data. This is an example of data leakage, as information from the test set (future data) leaks into the training phase.

To prevent this, the One-Hot Encoder was fitted only on the training data. During testing, the encoder applies the same transformation, ignoring any unseen categories (*San Francisco* in this case). This ensured the test data remains completely unseen during training, maintaining the integrity of the evaluation process.

- *Feature selection:* There are a variety of methods for feature selection. XGBoosts offers the Boosters objects where the get_score() method quantifies how useful each feature was in building the boosted decision trees within XGBoost model. Below are described different types of feature importance [134]:
  - 'weight': the number of times a variable is used to split the dataset across all trees.
  - 'gain': the average gain across all splits the variable is used in.
  - 'cover': the average coverage across all splits the variable is used in.

  This analysis used the "gain" feature importance, which measures a feature's contribution to reducing loss when used in a split. However, with one-hot encoding, each category becomes a separate column, spreading the importance across these columns and complicating interpretation. To address this, the importance scores for one-hot encoded columns were grouped and were summed to reflect the original categorical feature's overall importance. This was applied only to the dataset's original features, leaving the engineered features unchanged to preserve their individual contribution to the model's performance.
- *Model evaluation techniques* are essential for assessing the performance, reliability and generalization capabilities of a model [135]. Practices used include:
  - *TimeSeries Split:* In time series data, maintaining temporal order is crucial to prevent data leakage, ensuring past data is used to predict future data. For this analysis, the TimeSeriesSplit function from scikit-learn was used during cross-validation to generate train/test indices. This method splits the data into fixed time intervals, where test indices always follow training indices, preserving temporal order. Unlike standard cross-validation, which shuffles data, TimeSeriesSplit returns the first k folds as the training set and the (k+1)th fold as the test set in each split [136].

- *GridSearch:* An exhaustive search is conducted over a specified parameter grid for a model. This method from scikit-learn employs the cross-validation grid-search over a parameter grid [137]. However, in this analysis the TimeSeries cross validation was employed not to break the chronological order of the data. The scoring metric used in gridsearch was the "neg_root_mean_squared_error", which corresponds to the negative of the Root Mean Squared Error (RMSE). This approach optimized for lower RMSE values while identifying the best combination of hyperparameters. The parameter grid used was the following:

Hyperparameters	**Description**	**Values**
n_estimators	Number of gradient boosted trees.	[200, 300, 400, 500, 600]
max_depth	Maximum tree depth for base learners	[7, 9, 12, 15]
learning_rate	Boosting learning rate	[0.01, 0.1, 0.2]

Table 8: Hyperparameter grid applied to all features (XGBoost Documentation [Python API Reference], n.d.)

- *Regression Metrics:* The metrics used to assess the performance of the model in predicting order quantities are the following [138]**,** [139]**:**
  - Mean Absolute Error (MAE) measures the average magnitude of errors by calculating the absolute differences between predicted and actual values. Unlike squared error metrics, it is robust to outliers, as it avoids amplifying large errors.

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}^{-1}} |y_i - \hat{y}_i|$$

$$\hat{y}_i \rightarrow the\ predicted\ value\ of\ the\ i-th\ sample$$

$$yi \rightarrow\ actual\ value$$

$$n \rightarrow number\ of\ samples$$

- Mean Absolute Percentage Error (MAPE) measures the average absolute difference between predicted and actual values as a percentage, offering a scale-independent and easily interpretable metric, especially for non-technical stakeholders.

$$MAPE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}^{-1}} \frac{|y_i - \hat{y}_i|}{\max\,(\in, |y_i|)}$$

$$\in \rightarrow an\ arbitrary\ small\ yet\ positive\ number\ to\ avoid$$
$$undefined\ resultsin\ case\ y\ is\ zero$$

- Root Mean Squared Error (RMSE) calculates the square root of the squared differences between predicted and actual values, providing an error measure in the same units as the target variable. Scale-dependent, it offers intuitive insights into error magnitude while moderately penalizing large errors.

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}^{-1}} (y_i - \hat{y}_i)^2}$$

$$\hat{y}_i \rightarrow the\ predicted\ value\ of\ the\ i-th\ sample$$
$$yi \rightarrow\ actual\ value$$
$$n \rightarrow number\ of\ samples$$

- The coefficient of determination (R^2) measures the proportion of variance in the dependent variable explained by the independent features, indicating the model's goodness of fit. It ranges from 0 to 1, where 1 signifies perfect prediction and 0 indicates no explanatory power. Negative values may occur, reflecting a model that performs worse than a simple mean-based prediction. R^2 is scale-independent and easily interpretable, making it accessible to non-technical stakeholders.

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\hat{y}_i \rightarrow the\ predicted\ value\ of\ the\ i-th\ sample$$
$$yi \rightarrow\ actual\ value$$

$$n \rightarrow number\ of\ samples$$

$$where\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\ and\ \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i^2$$

Papers using the same dataset employed the following evaluation metrics: [87] used RMSLE, RMSE, MAPE and MAE; [118] used MAPE; [119] used RMSLE and MAE; and [113] used MAE, RMSE and MAPE.

For the first and second approach the following steps were applied:

1. The dataset was split chronologically, reserving the last ten weeks as the test set and the rest for training.
2. Categorical features were converted to numerical representations using One-Hot Encoding.
3. Data was divided into X_train, y_train (for training) and X_test, y_test (for testing).
4. A three-fold Time Series Split was used for cross-validation.
5. The XGBoost model was tuned using GridSearchCV with the previously defined hyperparameter grid.
6. Model performance was evaluated using MAE, MAPE, RMSE, and R^2 metrics on the test set.
7. A feature importance plot was generated, and the model was tested on the trimmed dataset using the optimal hyperparameters.

For the third approach the steps from one to five were the same. However, additional configurations were introduced due to the use of five lag features throughout the evaluation process.

Lag Features

According to [140] lag features, derived from past target values, play a critical role in time series modelling by capturing serial dependencies. For example, past order quantities can serve as input features to predict current. The number of lags to include depends on the dataset, as each lag provides different levels of information, though excessive lags may lead to redundancy.

Cycles and Lags

Lag features are particularly useful for capturing non-seasonal cycles (growths and decays) in the data that depend on prior values. Unlike single time series problems where ACF and PACF plots can guide lag selection, the dataset's structure (multiple meals across various centers) requires creating lag features for each group individually.

Although, calculating lag features is straightforward: the target variable is shifted by the desired lag, in this case the group function was used to separate individual meal-center combinations and prevent data leakage (using future data to predict the past). With the creation of lag features NaN values appear which can be handled efficiently by XGBoost.

Key Structure for Lag Features

A dictionary was initialized to store lag feature values dynamically. The keys in this dictionary were a combination of center_id, meal_id, and the lag number. This ensured that each lag feature was accurately linked to its specific meal-center combination and iteration step.

Dynamic Calculation of Lag Features in the Prediction Loop

- *Initialization:* For the first week of the test set, all lag features were initialized to zero since no prior predictions were available.
- *Dynamic Updates:* As the model iterated over each row in the test set:
  - For a given test row, the five lag features were dynamically calculated by retrieving the corresponding predictions from the dictionary. For instance, lag1 used the last prediction for the same center_id and meal_id, while lag2 referred to the prediction made two steps earlier, and so on for lag3, lag4, and lag5.
  - If no prior prediction was available for a specific lag, it defaulted to zero for that iteration.

Prediction and Dictionary Update

- The lag features, dynamically calculated for the current row, were added as input attributes to the model.

- The XGBoost model, trained with optimal hyperparameters from GridSearchCV, made a prediction for the current test row.
- After the prediction, the predicted value was dynamically added to the dictionary under the appropriate key (center_id, meal_id, lag_number). This ensured that future iterations used the most recent predictions for calculating lag features.

Evaluation and Feature Importance

Once the prediction loop was completed, the model was evaluated against the actual test values using metrics such as MAE, MAPE, RMSE, and R^2. A feature importance plot was created to assess the contribution of different features, including the lag features. Finally, the model's performance was tested on a refined dataset with non-significant features removed, using the same optimal hyperparameters from hyperparameter tuning.

This dynamic and iterative handling of lag features ensured accurate predictions while mimicking real-world production scenarios where only past predictions, not actual values, would be available for future forecasting.

4.1.3 XGBoost – Most Suitable Algorithm

Various models are available for regression tasks with tabular data. This thesis employed XGBoost (Extreme Gradient Boosting), a widely recognized model.

What is XGBoost?

XGBoost is an optimized, open-source, distributed gradient boosting library built within the Gradient Boosting framework [141], [142], [143], [144], [145]. Widely regarded as one of the most popular machine learning algorithms in recent years [141], [142], [143], it has helped individuals and teams excel in numerous machine learning and data mining competitions, particularly on Kaggle.

[142] noted that in 2015, 17 out of 29 Kaggle challenge-winning solutions utilized XGBoost, either alone or in combination with neural networks using ensemble techniques. In comparison, deep neural networks were used in 11 solutions. XGBoost

supports a range of tasks, including regression, classification, and ranking, and has been applied to problems such as store sales prediction and customer behavior modeling.

To understand how XGBoost operates, it is essential to explain several key concepts, including supervised machine learning, decision trees, ensemble learning, boosting, and gradient boosting [141], [142], [143], [144]:

- *Supervised Machine Learning:* In supervised learning, models are trained to identify patterns within a dataset containing features and a target variable. Once trained, the model predicts the target variable for new, unseen data.
- *Decision Trees:* They use a hierarchical structure where internal nodes represent tests on attributes, branches indicate test outcomes, and leaf nodes represent class labels. For example, a decision tree could predict house prices (labels) based on features such as the number of bedrooms and size. The tree learns by recursively splitting data into subsets based on attribute tests, stopping when the leaves are pure (all contain the same label) or when further splits do not improve predictions.
- *Ensemble Learning:* It combines multiple machine learning models to create a stronger, more accurate predictive model.
- *Boosting:* It is an ensemble method that builds a strong classifier from multiple weak classifiers. Each weak model is trained sequentially to correct the errors of its predecessor. The process continues until the training data is predicted correctly or no further improvement can be made.
- *Gradient Boosting:* It extends boosting by iteratively training weak learners to minimize the residuals (differences between predicted and actual values). The residuals form a loss function that is minimized using a gradient descent algorithm. When decision trees are used as weak learners, the model is referred to as a Gradient Boosting Decision Tree (GBDT).

How XGBoost Works

XGBoost is an optimized implementation of the Gradient Boosting Decision Tree (GBDT) model. It builds decision trees sequentially, assigning weights to all independent variables at the start. These weights are adjusted based on the prediction outcomes—decreased for correct predictions and increased for incorrect ones. The individual classifiers are then combined (ensemble) to produce a final, strong prediction [141].

Advantages of XGBoost

High Performance and Efficiency

XGBoost is highly optimized for speed and computational efficiency, making it one of the fastest gradient boosting libraries. It consistently delivers high-quality results across various machine learning problems [141], [142], [144]

Handles Missing Data

XGBoost can automatically manage missing values using a sparsity-aware algorithm. Missing data is classified into a default direction, and the model learns the best direction during training. This capability is particularly valuable for real-world datasets where missing values are common [141], [142], [144]

Regularization

To control model complexity, XGBoost incorporates L1 (Lasso) and L2 (Ridge) regularization [141], [144]. These techniques penalize overly complex models, reducing overfitting. Additionally, cross-validation enhances generalization and stabilizes performance [117], [144].

Scalability

XGBoost efficiently handles large-scale datasets using parallel and distributed computing, along with a cache-aware prefetching algorithm to minimize runtime. It can process billions of examples with fewer resources, running up to 10 times faster than other frameworks [144].

Captures Non-Linear Relationships

XGBoost models complex, non-linear relationships between features and target variables effectively [144].

Handles Non-Normal and Skewed Datasets

XGBoost, relying on decision trees, does not require features or target variables to follow a specific distribution, making it robust to non-normality and skewness.

Insensitive to Feature Scaling

Since decision trees split data based on thresholds rather than distances, XGBoost is unaffected by the scale or distribution of features. Unlike distance-based algorithms, it does not require feature scaling.

Handles Outliers

By focusing on reducing local errors during tree splits, XGBoost is less influenced by outliers compared to linear models, making it well-suited for skewed datasets with extreme values.

Final justification of XGBoost usage

XGBoost is a relatively user-friendly algorithm, even for non-technical users, making it an ideal choice for this thesis. The primary goal is to develop a web application for demand forecasting and optimization. To ensure the application is accessible and user-friendly, it is essential to implement algorithms that are both effective and resource-efficient. XGBoost meets these criteria due to its robust performance, high speed, flexibility, explainability, and ease of integration into various applications. It allows both technical and non-technical users to generate accurate forecasts and gain insights into feature importance—all without requiring expensive resources, even for large datasets-key elements for real-world applications.

ACKNOWLEDGMENTS

This work draws on insights from several research papers identified through a Systematic Literature Review process, many of which use the same dataset but apply different methodologies. These include:

- "Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis" by [87]
- "Machine Learning-Based Food Demand Estimation for Restaurants" by [118]

- "A Stack-Based Ensemble Model with Explainability for Food Demand Forecasting" by [119]
- "Effective Food Demand Forecasting Using Machine Learning Algorithms" by [113]

The first and the second paper influenced this work regarding the development of lag features.

Additionally, various Kaggle notebooks address the same problem using the same dataset applying diverse approaches. Both effective and flawed practices from these notebooks have been critically evaluated and some of them are incorporated in this work.

Special mention is given to [140] for his three articles on time series forecasting for energy consumption using XGBoost and LightGBM. His approach incorporates additional features, lag features and other practices which some of them significantly enhanced model performance and have inspired aspects of this work.

4.1.4 Introduction of the dataset

The dataset utilized originates from a Kaggle competition titled “Food Demand Forecasting”. The company is a meal delivery service operating across multiple cities. To ensure efficient operations, it maintains several fulfillment centers responsible for dispatching meal orders to customers. The company seeks assistance in forecasting future demand to enable these centers to better plan their raw material inventory for the upcoming weeks.

The following files were available through notebooks in Kaggle except the test file which includes weeks 146-155:

Table 9: train file description

Attribute	Description
id	Unique identifier
week	Week number
center_id	Unique identifier of each distribution center

meal_id	Meal unique identifier
checkout_price	Price when sold
base_price	Original price
email_for_promotion	If an email was used for promotion
homepage_featured	If a meal was displayed in the homepage
num_orders	Number of orders for specific meal

Table 10: fulfilment_center_info file description

Attribute	**Description**
center_id	Distribution center unique identifier
city_code	City unique identifier
region_code	Region unique identifier
center_type	Distribution center type
op_area	Operation area

Table 11: meal_info description

Attribute	**Description**
meal_id	Meal unique identifier
category	Meal category
cuisine	Meal cuisine type

4.1.5 Exploratory Data Analysis (EDA)

At this point, the three datasets were merged to facilitate the coding process especially during the EDA stage.

First, the author looked for general information such as number of rows, number of columns, data types of each attribute and the absence of any data. The figure below shows that there are 456548 entries, 15 columns, integers, floats and objects (text) as data types and no missing values.

```
# Check the number of rows, columns and on-null values.
# Also, the types of data are taken into account
final_merged_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 456548 entries, 0 to 456547
Data columns (total 15 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   id                     456548 non-null  int64
 1   week                   456548 non-null  int64
 2   center_id              456548 non-null  int64
 3   meal_id                456548 non-null  int64
 4   checkout_price         456548 non-null  float64
 5   base_price             456548 non-null  float64
 6   emailer_for_promotion  456548 non-null  int64
 7   homepage_featured      456548 non-null  int64
 8   num_orders             456548 non-null  int64
 9   city_code              456548 non-null  int64
 10  region_code            456548 non-null  int64
 11  center_type            456548 non-null  object
 12  op_area                456548 non-null  float64
 13  category               456548 non-null  object
 14  cuisine                456548 non-null  object
dtypes: float64(3), int64(9), object(3)
memory usage: 52.2+ MB
```

Figure 5: Merged dataset information

As it was stated in the beginning of this chapter, the dataset includes a variety of meals distributed across different centers over multiple weeks. Therefore, there was no single time-dependent entity (e.g., one meal in one center over time) to visualize directly.

However, by grouping the data based on the "week" attribute, a consistent time series representation can be created only for demonstration purposes, allowing for a clear visualization of the total weekly number of orders. This can be seen below:

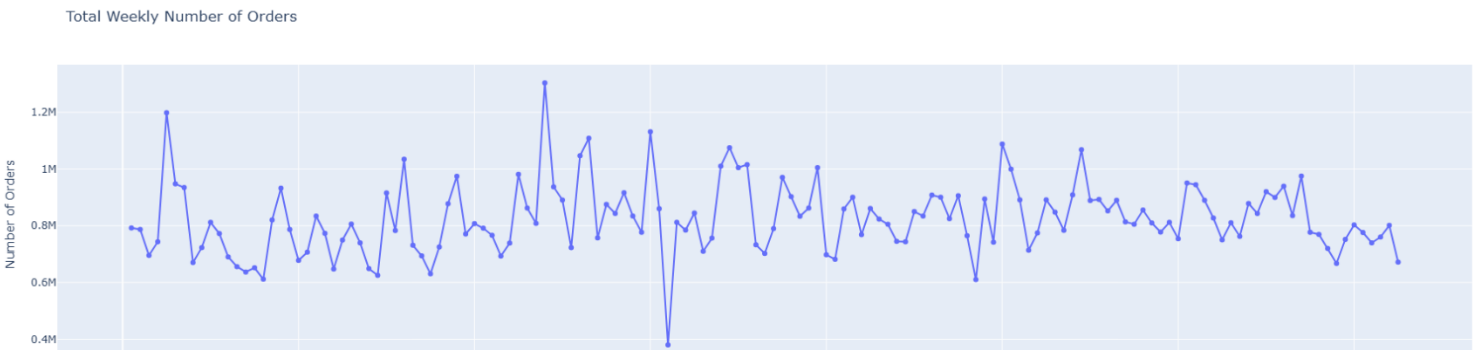

Figure 6: Total aggregated number of orders by week

The figure below shows both a simple linear trend and a seasonal component. One can dive into another time series by grouping the data by a specific meal and distribution center as shown below:

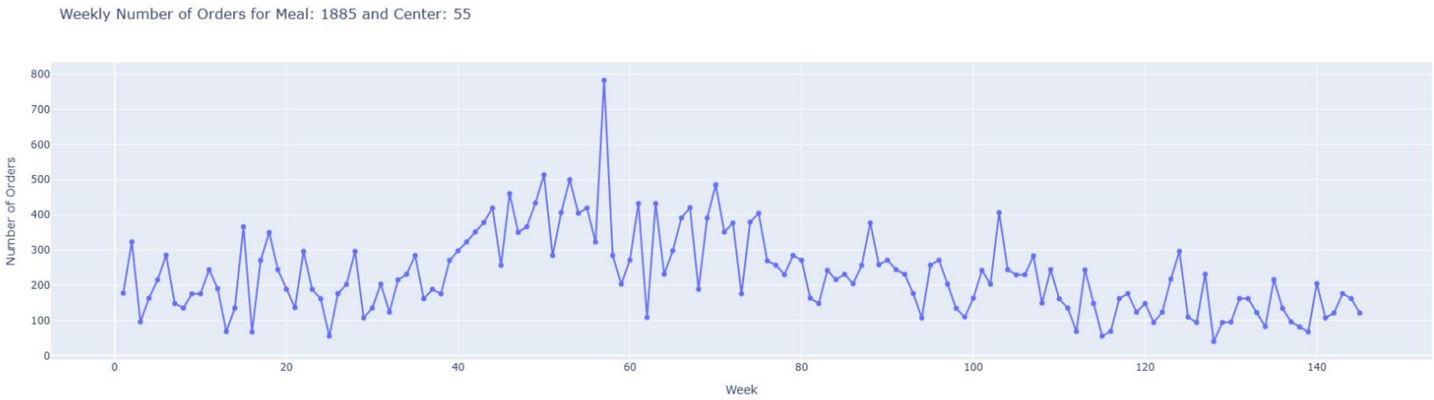

Figure 7: Aggregated number of order for meal 1885 and center 55

When analyzing datasets, it is crucial to examine the distribution of numerical features and identify outliers. Histograms and boxplots are common tools used for this purpose. In the merged dataset, three numerical attributes—checkout_price, base_price, and num_orders—are analyzed.

The histogram for the num_orders feature revealed high positive skewness, evident from the long tail on the right-hand side. This indicated the presence of a few significantly larger values compared to the rest of the data. Such skewness can arise from factors like special promotions, seasonal peaks, marketing campaigns, outliers, or unusual customer demand. Slightly positive skewness was observed in the other two price variables as well.

Boxplots effectively summarize feature distributions and highlight outliers as circular markers. While all three attributes contained outliers, their impact was most pronounced in the num_orders feature, contributing significantly to its skewed distribution.

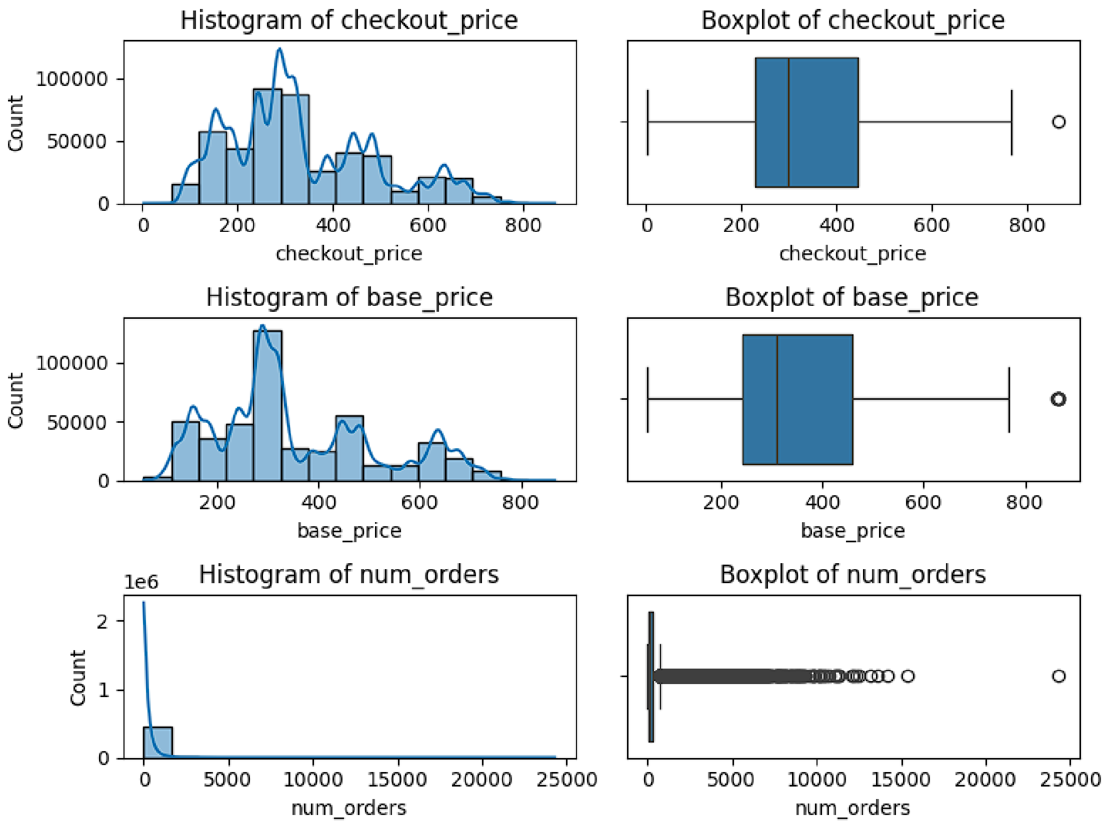

Figure 8: Histograms and Boxplots of numerical attributes

Another common practice when analyzing datasets is the exploration of any relationships between quantitative variables. This can be done using the Pearson and Spearman correlation coefficients. However, there are some key differences between these two methods that are described below [146]:

- The Pearson correlation is applied for linear relationships, whereas Spearman is valid for both linear and non-linear relationships.
- The Pearson correlation quantifies the relationship between attributes in terms of degree of similarity or difference. On the other hand, Spearman correlation measures the relationship between features in terms of degree of ordering.
- The Pearson correlation can be applied to only quantitative and continuous variables while the Spearman correlation can be computed also to ordinal and categorical features.
- The Pearson correlation is affected by extreme values, while the Spearman is less sensitive to outliers.

Both of these metrics range from -1 to 1 indicating a perfect negative relationship and a perfect positive relationship respectively. A value of 0 shows no relationship between the variables.

Both the Pearson and the Spearman correlation metrics were calculated using a function in Python which integrates the following formulas [147]:

Pearson Correlation Coefficient Formula

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

ρ: Pearson correlation coefficient

cov: covariance of variables x and y

σ_X: standard deviation of x

σ_Y: standard deviation of y

Spearman Correlation Coefficient Formula

$$r_s = 1 - \frac{cov[R[X], R[Y]]}{\sigma_{R[X]} \sigma_{R[Y]}}$$

cov[R[X], R[Y]]: covariance of the rank variables

$\sigma_{R[X]}$. $\sigma_{R[Y]}$: standard deviations of the rank variables

The below figure illustrates both the Pearson and the Spearman correlation matrices in a visual way. There is a strong positive relationship between checkout_price and base_price. Furthermore, a weak or moderately negative relationship between the num_orders and the two price variables is depicted too, where Spearman suggests slightly stronger relationships compared to Pearson, which could indicate that the relationship may not be perfectly linear but still monotonic a key component of the Spearman metric.

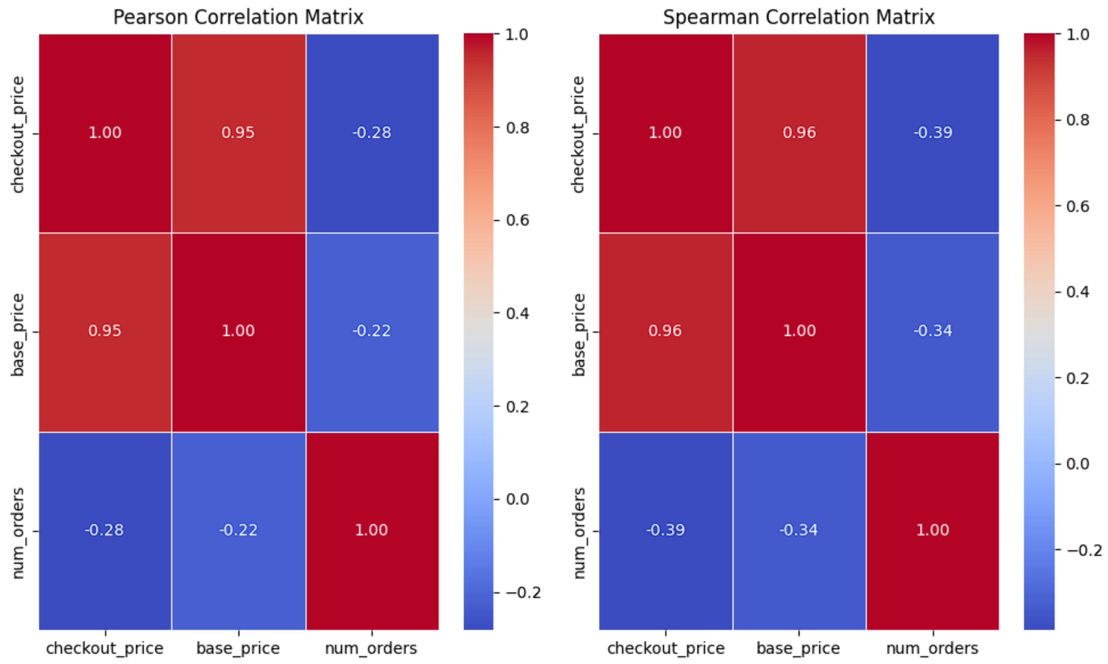

Figure 9: Pearson and Spearman correlation matrices

Another way of observing the relationships and distributions among quantitative variables is the well-known scatterplot matrix. The diagonal plots demonstrate the histograms, while all the others the relationships created between the features. The below figure supports the correlation matrices' findings since it indicates a strong positive relationship between checkout_price and base_price and the week negative relationships between num_orders and the two price features. Also, the outliers mainly found in the num_orders attribute cannot be ignored and, in many cases, they might need further investigation.

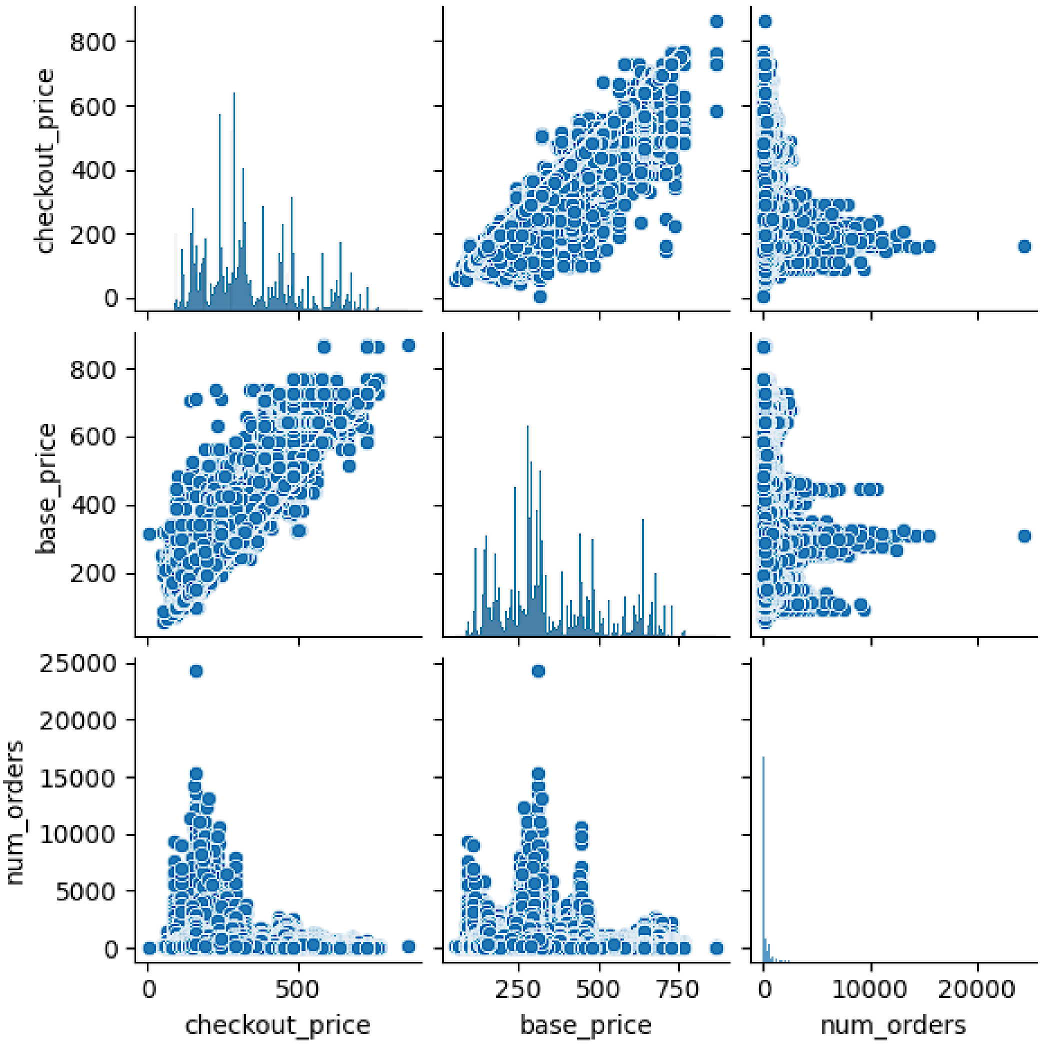

Figure 10: Scatterplot Matrix

There are several ways to analyze categorical variables, including plots, tables, and statistical tests. However, domain knowledge is crucial to uncover unique patterns in real-world datasets. In this case, a bar plot revealed that Italian cuisine was the most ordered type in the dataset, as shown in the figure below:

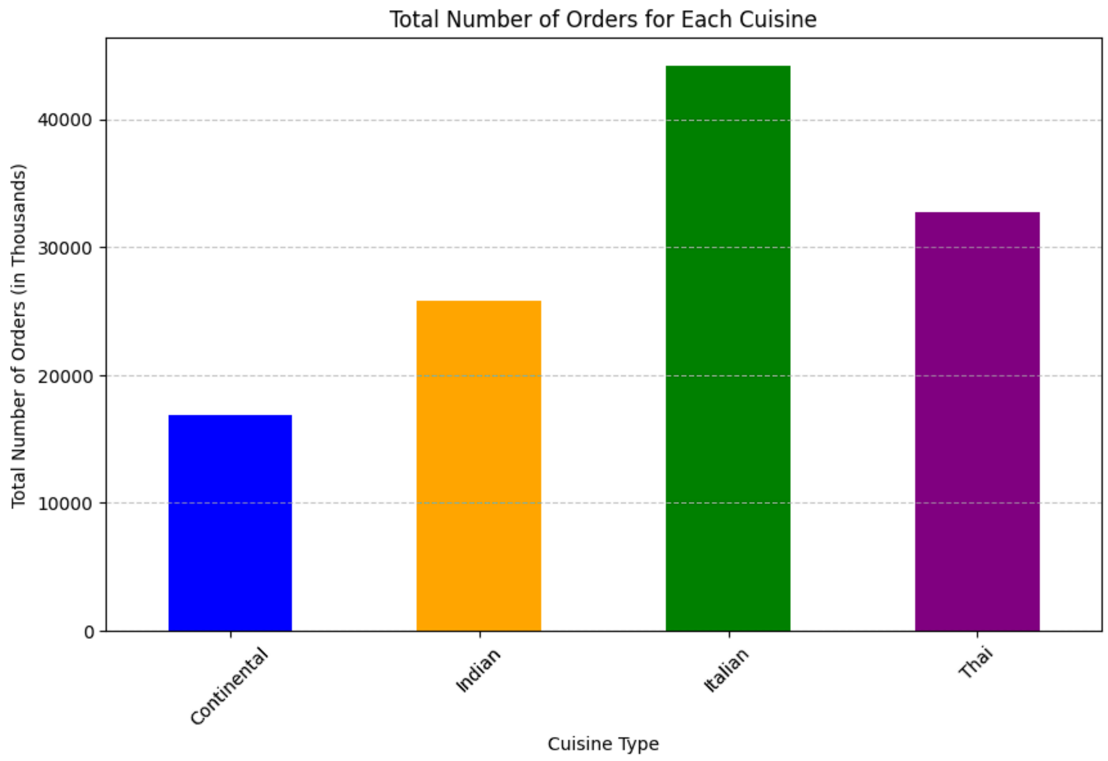

Figure 11: Total number of orders for each cuisine type

A comprehensive Exploratory Data Analysis (EDA) was conducted, providing a clear understanding of various aspects of the dataset, including distributions, outliers, missing values, and visualizations of categorical variables.

The next phase of the analysis involved implementing five distinct approaches to progressively enhance the model's performance. Throughout these experiments, preprocessing, feature engineering, and model evaluation techniques were applied iteratively to refine and assess the results. Additionally, as previously mentioned, the dataset does not contain a single time-dependent variable. Consequently, the problem was approached with machine learning methods while including time-series related practices.

4.1.6 1st attempt with all features

Using all available features from the three files, the optimal hyperparameters determined through GridSearchCV were: learning_rate=0.1, max_depth=7, and n_estimators=200. The model's performance was evaluated as follows:

Table 12: Evaluation Metrics Result with all features model

Metric	**All Features Model**
Mean Absolute Error (MAE)	76.99
Mean Absolute Percentage Error (MAPE)	66.71%
Root Mean Squared Error (RMSE)	126.29
R^2 Score	0.79

The results indicated that the model explained 79% of the variance in the target variable (R^2 = 0.79). The errors, as measured by RMSE and MAE, were relatively moderate, suggesting a reasonable prediction accuracy, though the MAPE of 66.71% highlighted potential challenges in capturing smaller values or high variability in the data.

The feature importance plot showed that the "center" feature contributed the most to reducing the loss function, while the "id" feature contributed the least. This highlighted that the "id" feature provided little to no meaningful information for identifying underlying patterns.

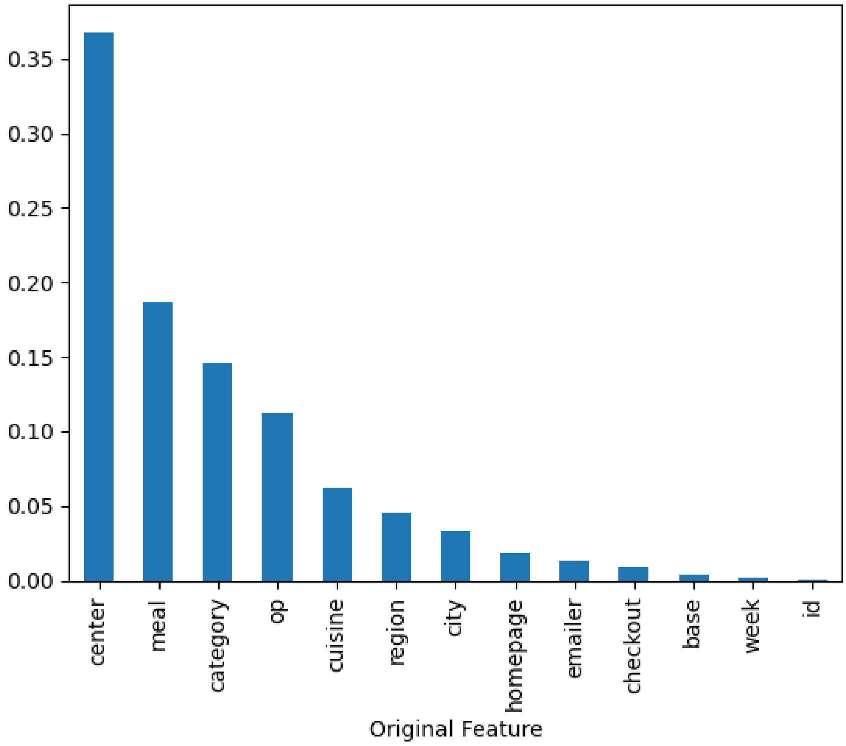

Figure 12: Feature importance plot using all features

Rerunning the model without the "id" attribute using the optimal hyperparameters did not improve performance. However, since "id" had minimal impact on reducing the loss function, it was excluded from the dataset.

4.1.7 2nd approach with additional features

In this section, additional features were engineered to assess their impact on the model's performance. These features can be seen below:

Dataset-wide features: Simple row-wise features were calculated such as meal profitability which is simple the subtraction between base_price and checkout_price. Since, the calculations were done row by row no future information was leaked into the training phase. Therefore, no data leakage existed in this scenario.

Dataset-wide features

The features created below were introduced in various Kaggle notebooks, including the following examples:

- "R^2 Score: 80% | 10-Week Food Demand Forecasting" by Ali Almulla
- "END-TO-END Food Demand Forecasting R2 0.90" by Tarık Yılmaz

The features included:

- NewDiscountRate → (base_price – checkout_price) / base_price
- NewMealProfitability → base_price – checkout_price
- NewCategoryCenter → category + center_type
- NewCuisineCenter → cuisine + center_type

The new dataset, employed by the additional engineered features and without the non-important “id” attribute, throughout the GridSearch approach provided the optimal hyperparameters as 'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 200. The evaluation metrics resulted are the following:

Table 13: Evaluation metrics using additional features

Metric	**All Features**	**Additional Features**	**Improvement**
Mean Absolute Error (MAE)	76.99	77.19	↑ 0.20 (0.26% increase)
Mean Absolute Percentage Error (MAPE)	66.71%	66%	↓ 0.71% (1.06% reduction)
Root Mean Squared Error (RMSE)	126.29	125.56	↑ 0.73 (0.58% reduction)
R^2 Score	0.79	0.79	0.00 No change

The results indicated that introducing additional features led to a slight improvement in MAPE (-1.06%) and RMSE (-0.58%), suggesting better prediction accuracy and reduced error magnitude. However, MAE increased slightly (+0.26%), indicating a marginal rise in average absolute error. The R^2 score remained unchanged, implying that the model's overall explanatory power did not improve significantly. These results suggested that the new features had a minimal but nuanced impact, primarily benefiting percentage-based metrics like MAPE.

The plot below illustrated that two of the four newly created attributes such as “New-MealProfitability” and “NewMealProfitability” did not contribute at all to the reduction of the loss function. Therefore, they will be excluded for future model development.

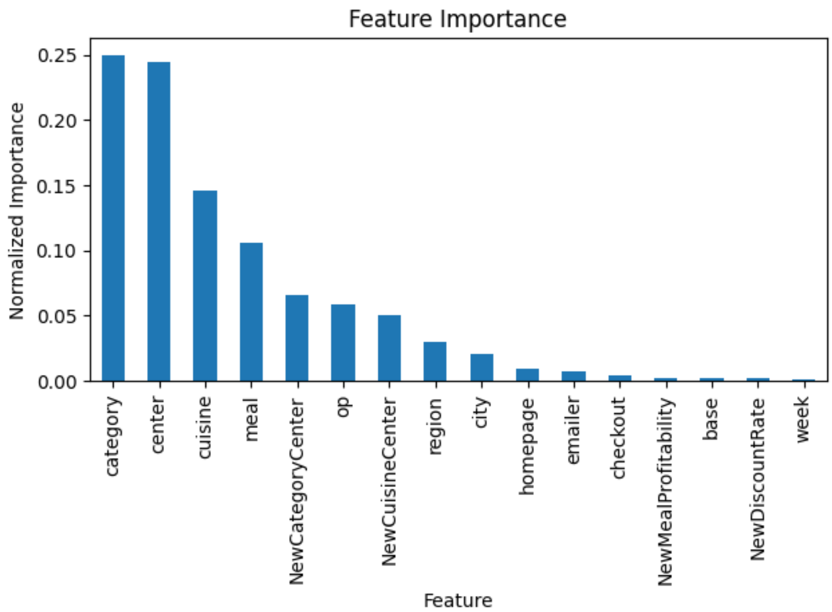

Figure 13: Feature importance plot using additional features

The model ran once more without the “NewMealProfitability” and “NewMealProfitability” resulting in slightly better performance with evaluation metrics as MAE 76.20, 65.78%, 123.98 and R^2 0.80.

4.1.8 3rd attempt with lag features

Employing the idea of five features and running the model within GridSearch method the optimal hyperparameters resulted as 'learning_rate': 0.01, 'max_depth': 7, 'n_estimators': 300. The evaluation results table is shown below:

Table 14: Evaluation metrics using lag features

Metric	Additional Features	With Lag Features	Change
Mean Absolute Error (MAE)	77.19	171.44	↑ 94.25 (122.12% increase)
Mean Absolute Percentage Error (MAPE)	66%	74.28%	↑ 8.28 (12.55% increase)
Root Mean Squared Error (RMSE)	126.56	311.16	↑ 184.60 (145.92% increase)
R^2 Score	0.79	-0.27	↑-1.06 (134.18% increase)

The table displays the non-positive effect of lag features in model performance. The MAE, RMSE and R^2 score increased significantly, indicating the accuracy was very low. Specifically, the R^2 score highlighted that the model performed worse than a baseline model that simply predicts the mean of the target variable for all data points. The MAPE showed the relative error increased by almost 13%, suggesting a decreased accuracy too. The addition of lag features decreased the model's ability to capture historical patterns, resulting in very low prediction accuracy and a worse overall fit. This indicated that time-dependent information significantly did not contribute to understanding and predicting the target variable.

From the feature importance plot below, it is observed that the num_orders_lag_1 feature had the highest importance, indicating that the number of orders from one lagged period was the most influential predictor of the target variable. Other lagged features (num_orders_lag2, num_orders_lag3, etc.) did not contribute a lot. In conclusion, the influence of lag features was not so strong highlighting the unimportance of historical data for forecasting. Only the lag one feature remained for future model development.

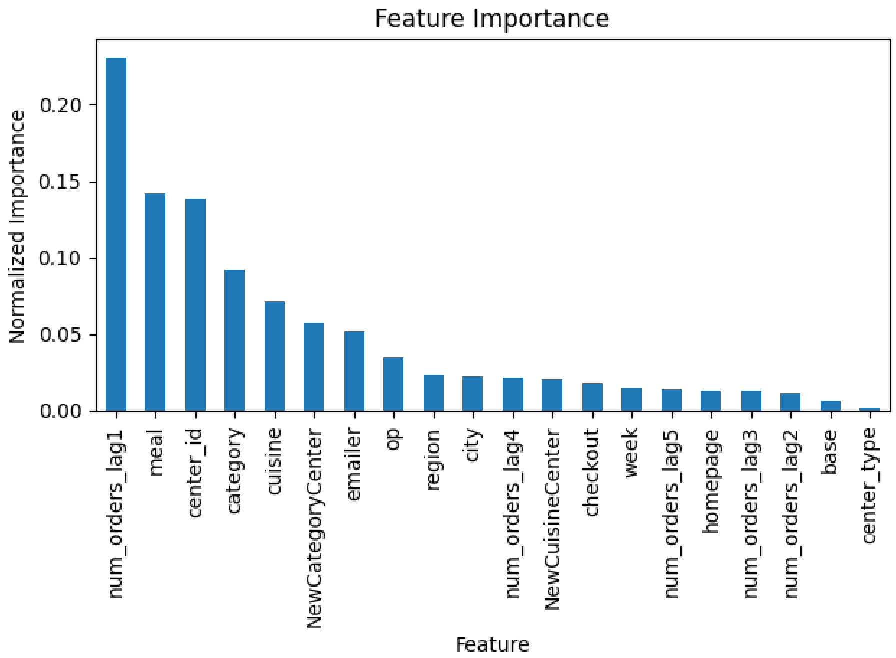

Figure 14: Feature importance using lag features

The model was re-evaluated using only the lag-1 feature, yielding improved performance compared to using all five lag features. However, its performance remained inferior to the model with additional features alone. The evaluation metrics were: MAE 132.47, MAPE 97.44%, RMSE 245.91, and R^2 score 0.21. Consequently, lag features were excluded from the model development process.

In conclusion, the best results provided by the model using additional features such as “NewCategoryCenter” and “NewCuisineCenter”. This predicted demand will be used later for further development within an optimization model.

4.2 Linear Optimization Method

As was stated during the Systematic Literature Review process, inventory management is a critical process in the food industry. With the growing population, the demand for efficient inventory management systems continues to rise to ensure a steady supply of stocked food items. A traditional technique in inventory management is the Economic Order Quantity (EOQ) model, which helps companies determine the optimal order quantity to meet their needs efficiently. However, this approach operates under simplified assumptions, such as constant demand and the absence of constraints, making it less applicable in real-world scenarios that involve capacity, budgetary, or logistical limitations.

To address these limitations, a modified version of EOQ, such as the EOQ Lagrange method introduced by [124], incorporates one or two constraints to determine optimal lot sizes. While this represents a step forward, the complexity and dynamic nature of crisis-driven food supply chains require a more adaptable and robust solution. In our case, a linear optimization method was utilized to minimize the total cost and determine the optimal order quantities for different meals across distribution centers over a ten-week planning horizon.

This technique provides high flexibility as it can integrate multiple constraints, such as capacity limitations and budget restrictions, into the optimization process. The optimization model considers the predicted demand generated by the XGBoost forecasting algorithm and calculates the optimal ordering strategy to meet this demand while adhering to the specified constraints. This approach ensures efficient resource allocation, reduces waste, and improves the overall resilience of the supply chain during crisis situations.

Furthermore, a cost reduction sensitivity analysis and a demand simulation were conducted to further enhance the user's ability to gather insightful information.

4.2.1 Application of Linear Programming

Methodology for Main Model

This methodology outlines the steps taken to develop and solve an integer programming optimization model for multi-center, multi-meal inventory and order planning. The goal was to minimize overall costs while considering constraints such as budgets, warehouse capacities, and demand satisfaction.

Data Preprocessing

1. *Data Input:* The predicted demand resulted from the XGBoost model was used alongside other information such as centers, meals, and checkout prices. The period concerned the last ten weeks of each combination of center and meal.
2. *Filtering:* Users selected specific centers and meals for inclusion in the model, and corresponding data was extracted.
3. *Parameter Collection:* Key parameters were gathered, including ordering costs, holding costs, and unit costs for each selected meal. Weekly budgets and warehouse capacities were specified for each center, along with initial inventory levels for each center-meal combination during the first week.

Baseline Budget Estimation

- A baseline budget was estimated for each center by summing the expected weekly costs of fulfilling predicted demand of all selected meals with a safety margin of 20%. This budget calculation was used as a baseline suggestion for the user as an initial starting point to cover all the demand.

Optimization Model Formulation

1. *Objective Function:* Minimize total costs, comprising:
    - Ordering costs (fixed and variable),
    - Inventory holding costs,
    - Penalty costs for unmet demand.

$$Z = \sum_{c \in C} \sum_{m \in M} \sum_{t \in T} (o_m I_{c,m,t} + u_m Q_{c,m,t} + h_m S_{c,m,t} + p_{c,m,t-1} U_{c,m,t})$$

Where:

- C: Set of Centers
- M: Set of Meals
- T: Set of Weeks
- o_m: Ordering cost for meal m
- u_m: Unit cost for meal m
- h_m: Holding cost for meal m
- $p_{c,m,t}$: Checkout price for meal m in week t at center c
- $I_{c,m,t}$: Binary variable (1 if meal m is ordered at center c in week t, 0 otherwise)
- $Q_{c,m,t}$: Quantity of meal m ordered at center c in week t
- $S_{c,m,t}$: Inventory of meal m at center c in week t
- $U_{c,m,t}$: Unment demand for meal m at center c in week t

Note

The Checkout price storage structure in Python was implemented as a dictionary with zero-based indexing. Consequently, the index i-1 was used to correctly reference the corresponding demand values.

2. *Decision Variables:*
    - Order quantities, inventory levels, and unmet demand (continuous variables),
    - Order indicators (binary variables) indicating if an order is placed with the 1 or 0 otherwise.
3. *Constraints:*
    - *Inventory Balance:* Ensured inventory consistency across weeks considering demand, lead times, and initial inventory.

      For t =1:

      $$S_{c,m,t} = S_{c,m,0} - D_{c,m,t-1} + U_{c,m,1}$$

      Else:

      $$S_{c,m,t} = S_{c,m,t-1} - Q_{c,m,t-lead\ time} - D_{c,m,t-1} + U_{c,m,t} - U_{c,m,t-1}$$

      Where:

      $D_{c,m,t}$: Predicted demand for meal m at center c in week t lead time: the time it takes from the placement of an order until the actual delivery. It is set to be 1.

For week 1, the inventory was calculated based on the demand for that week. If the inventory was insufficient to meet the demand, the unmet demand variable increased accordingly. Conversely, if the inventory exceeded the demand, the surplus was carried forward.
For weeks greater than 1, the inventory calculation incorporated the order quantity and the unmet demand from the previous week. The first term accounted for the lead time, determining when the order will be received (i.e., 1 week after), while the second term ensured that the previous unmet demand was factored into the inventory balance.

Note

The demand storage structure in Python was implemented as a dictionary with zero-based indexing. Consequently, the index i-1 was used to correctly reference the corresponding demand values.

- *Budget Limit:* Total spending per week (fixed and variable) for all the selected meals was limited by user-defined budgets for each center.
  The total ordering and holding costs for all meals must not exceed the weekly budget for each center:
  $$\sum_{m \epsilon M} (o_m \, I_{c,m,t} + u_m Q_{c,m,t}) \ \leq B_c$$
  where B_c is the weekly budget for center c
- *Warehouse Capacity:* Restricted the total inventory at each center for all the selected meals to its storage capacity.
  The total inventory of all meals for each center must not exceed its weekly warehouse capacity:
  $$\sum_{m \epsilon M} (o_m \, I_{c,m,t} + u_m Q_{c,m,t}) \ \leq W_c$$
  where W_c is the warehouse capacity for center c.

- *Linking Binary Variable to Order Quantity:* Binary variables linked to order quantities using a Big-M approach.
  To ensure the binary variable $I_{c,m,t}$ correctly indicates whether a meal was ordered, the order quantity must satisfy:
  $$Q_{c,m,t} \leq M I_{c,m,t}$$
  In our case M was a very large number i.e., 1000000
- *Non-Negativity:*
  All decision variables must be non-negative:
  $$Q_{c,m,t} S_{c,m,t} U_{c,m,t} \geq 0 \ \forall c \in C, m \in M, t \in T$$

Model Implementation and Solution

- The optimization problem was implemented using Python and solved using the PuLP linear programming library [148]. Decision variables were optimized to satisfy the constraints while minimizing the objective function.

Output and Metrics

- *Optimization Results:* Generated outputs for weekly order quantities, inventory levels, and unmet demand for each center and meal.

Performance Metrics

- Total costs (ordering, holding, and penalty costs),
- Total unmet demand and its frequency,
- Budget utilization by center.

This methodology provided a scalable and flexible framework to optimize supply chain operations across multiple distribution centers and product categories.

Input for Main Model

The study involved two distribution centers, labeled as Center 55 and Center 24, and two types of meals, identified as Meal 1885 and Meal 1993.

Distribution Centers and Warehouse Details

- The warehouse capacity for both centers is 1,000 units.
- The initial inventory for all center-meal combinations in the first week was set to 0.

Meal Details

- *Meal 1885:*
  - Ordering Cost: 100
  - Holding Cost: 10
  - Unit Cost: 10
- *Meal 1993:*
  - Ordering Cost: 500
  - Holding Cost: 5
  - Unit Cost: 5

Budget Information

- For Center 55, the suggested budget was approximately 4,781, while the researcher specified a budget of 4,500.
- For Center 24, the suggested budget was around 21,389, with the researcher setting the budget at 21,000.

Performance metrics and visualization plot for main model

The model resulted in a total cost of approximately 980.000 with the majority attributed to penalty costs incurred from unmet weekly demand. Penalty costs were inevitable in the first week due to a lead time of 1 and the decision to start with 0 initial inventory. Despite this, the unmet demand for each meal was fully addressed by the end, as the baseline budget suggestions were used to cover all demand. Budget utilization was around 98% for Center 55 and 88% for Center 24, as shown below:

Table 15: Performance Metrics for main model

Total Cost	**Ordering Cost**	**Holding Cost**	**Penalty Cost**	**Total Unmet Demand by the end period**	**Unmet Demand Frequency**	**Budget Utilization**
979404.25	205588.17	1555.60	772260.49	0.00	14 weeks	Center 55: 98% Center 24: 88%

To get a better picture of how the order quantity evolves over time alongside any unmet demand existing the following plot was created:

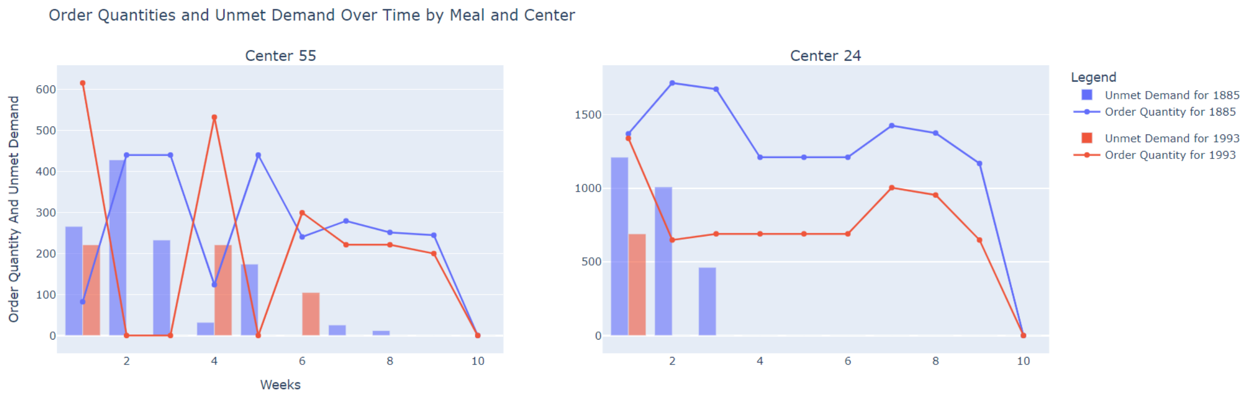

Figure 15: Order Quantities and Unmet Demand Over Time for model first run

From the figure above, it is observed that the demand was higher in center 24 especially for meal 1885. Furthermore, both centers experience significant unmet demand in the first 2 weeks, particularly for meal 1885 in center 24. Also, it is observed that the order quantity at week 10 is 0 because there was no demand to be fulfilled at week 11.

4.2.2 Cost Reduction Sensitivity Analysis

Sensitivity analysis is the study of how uncertainty in the output of a mathematical model can be attributed to variations in its inputs. It involves quantifying the influence of individual or combined inputs on the model's results, often through sensitivity indices that measure these effects [149]. In other words, it explains "how" and "how much" alterations in the parameters of an optimization model influence the objective function and the optimum point [150]. Today, additional knowledge is needed apart from just running a model and getting an answer. It is significant to evaluate the effects of changes in input parameters and to provide an understanding of how the model responds to these alterations [150].

In the context of budget reduction, sensitivity analysis evaluates how varying budget levels affect outcomes such as costs, inventory, and unmet demand, providing insights into the model's robustness and identifying critical input thresholds that significantly impact performance.

Methodology and Input for Sensitivity Analysis

Budget Reduction Levels:

Budget reduction percentages were provided as input (e.g., 10, 20). These percentages were converted into decimal values and used to scale down the original budgets for each center. In this case, the researcher applied five budget reduction policies which are 10%, 20%, 30%, 40% and 50%.

The formula of budget reduction is:

$$Adjusted\ Budget_{center} = Budget_{center}\ (1 - \frac{Reduction\ Percentage}{100})$$

This allows the model to test the feasibility of meeting demands under increasingly constrained financial conditions.

Sensitivity Analysis Results

From the below table it is illustrated that reducing the budget the total cost significantly increases, driven primarily by the penalty cost. This cost corresponds to increasing unmet demand levels. The holding cost remains in low values throughout, suggesting that inventory is not retained for long periods.

Table 16: Performance Metrics for Budget Reduction Sensitivity Analysis

Budget Reduction %	**Total Cost**	**Ordering Cost**	**Holding Cost**	**Penalty Cost**	**Total Unmet Demand by the end period**	**Unmet Demand Frequency**	**Budget Utilization**
10	1.485542e+06	202274.22895	4193.599910	1.279074e+06	211.39394	22	Center 55: 100% Center 24: 97%
20	3.065401e+06	183599.99900	3017.009158	2.878784e+06	2018.81694	27	Center 55: 100% Center 24: 100%
30	4.799218e+06	160649.99950	1401.175980	4.637167e+06	4363.81690	27	Center 55: 100% Center 24: 100%
40	6.529239e+06	137699.99993	1105.057070	6.390434e+06	6648.81690	27	Center 55: 100% Center 24: 100%
50	8.277060e+06	114697.47145	4283.555595	8.158079e+06	8919.06980	28	Center 55: 100% Center 24: 100%

The figure below illustrates in a graphical way the table results described earlier. While the ordering cost decreases due to budget constraints the unmet demand increases. The holding cost is observed to increase mainly at 10% and 50% budget reduction.

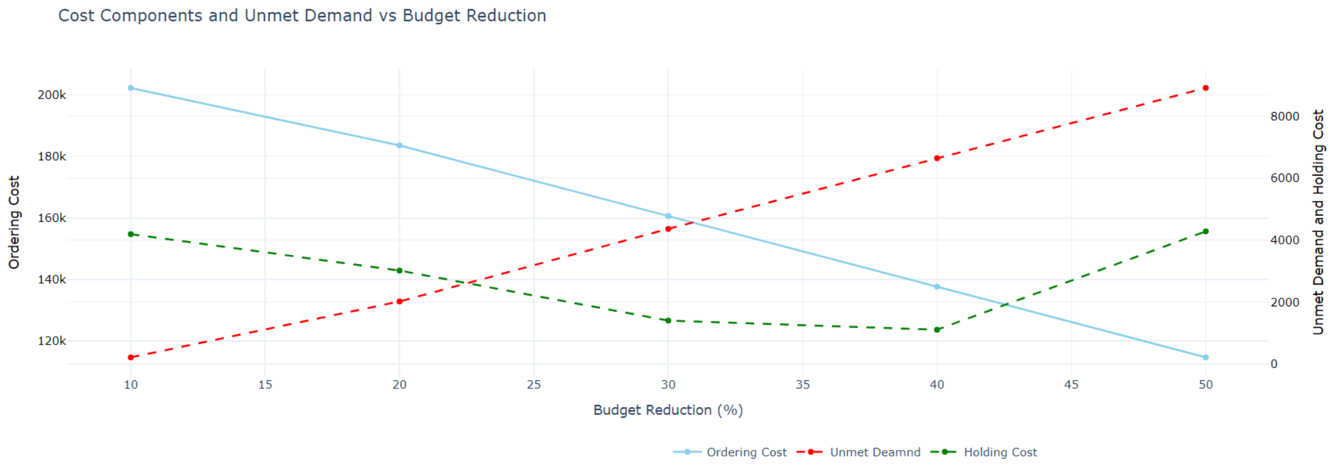

Figure 16: Budget Reduction Sensitivity Analysis Plot

4.2.3 Demand Variation Simulation

A simulation is a representation that replicates the behaviour of a real or planned system, enabling the testing of various scenarios or adjustments to processes to support informed decision-making [151]. Computer simulation is used for modelling real-world situations or hypotheses on a computer so that it can be study further to see how the system operates. By altering the variables in the simulation, conclusions can be drawn about the behaviour of the system [152].

In the context of demand variation, simulation allows us to assess how changes in demand levels influence the outcomes of an optimization model, providing insights into the robustness and adaptability of the system.

Demand Variation Method

Application of random adjustments to the initial demand values was conducted by multiplying a factor by a uniform distribution within a specified range. This range was given by the user for example a scenario could be that the demand can vary between -10% and +10%. This generated new demand values that reflect fluctuations and allows for an assessment of how the optimization model responds to changes in demand under different scenarios.

$$D_{new} = D_{original}\,(1 + \epsilon)$$

Where:

D_{new} is the new demand after variation

$D_{original}$ is the original (base) demand value

ϵ is the random variation factor and is drawn from a uniform distribution in a specified range (-scenario, +scenario)

In this case, the input was 0, 0.1,0.2,0.3,0.4 and 0.5 percent variation of demand. Each of these five scenarios ran for 100 simulations.

Demand Simulation Results

The results cannot be displayed since each of the five scenarios ran for 100 simulations resulting in 600 rows.

To make usage of the results the following line plot was created averaging the unmet demand frequency across the different scenarios. The unmet demand frequency rises steadily across scenarios as demand variations increase. More specifically, the unmet demand frequency shows a sharper increase in higher variation scenarios while the biggest value is observed at 40%, indicating greater sensitivity to larger fluctuations in demand. Overall, the increase in unmet demand frequency is relatively mild, suggesting the system can handle minor demand variations effectively.

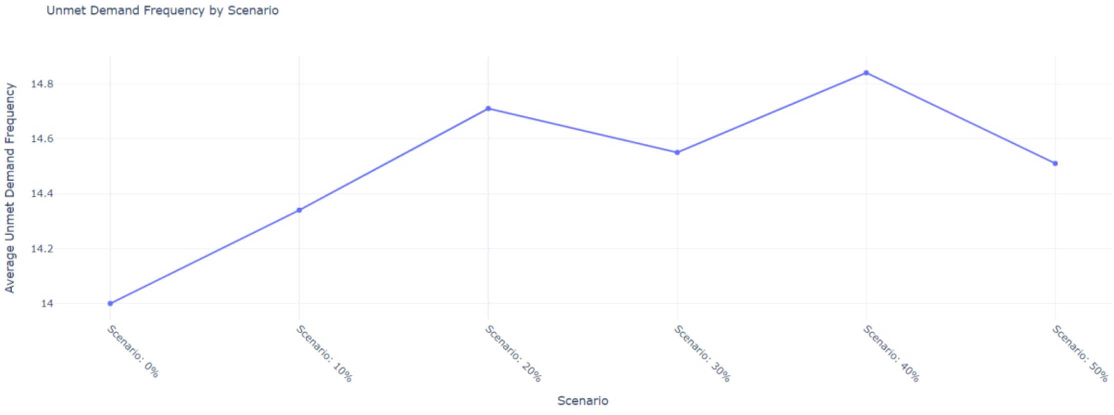

Figure 17: Demand Variation Simulation Plot

4.3 Introduction to Web Application

The web application was designed to provide users with an accessible, interactive, and user-friendly interface for utilizing predictive and optimization models. It was developed using Streamlit, an open-source Python framework that allows for the rapid creation of interactive data apps with minimal code [153]. Additionally, Streamlit offers the option to deploy apps publicly via the Streamlit Community Cloud [154]. Users can easily create a free account, link their GitHub repository containing the app (app.py) and the requirements.txt file, and deploy the app online. However, the free version of Streamlit Community Cloud offers limited memory (1GB), no GPU, and restricted runtime, which proved insufficient for handling the large dataset required during the machine learning and optimization phases. As a result, the web app was not made publicly available in this thesis. By using the google colab paid version we can enhance both the GPU and the memory capacity in order to support the implementation of the streamlit app.

Ngrok was used to expose the locally running Streamlit application on Google Colab to the public internet [155]. Since Google Colab does not provide a native way to expose local applications to the web, ngrok was employed to create a secure tunnel to the Colab environment, generating a temporary public URL for accessing the Streamlit app. This allowed for easy remote access and testing without the need for permanent hosting. This approach was particularly useful for development and sharing purposes, as the free Streamlit Community Cloud platform had memory limitations that hindered deployment for large datasets.

Web Application Interface

The web application is organized into four main sections: uploading datasets, performing Exploratory Data Analysis (EDA), predicting order quantities using the XGBoost model, and running the optimization model. Each of these phases is interactive, providing the user with a smooth and guided experience. The application displays results such as tables and plots that are fully interactive, allowing the user to gain detailed insights at each step.

Uploading Datasets

The user can upload three essential CSV files: train, fulfilment_center_info, and meal_info, as shown in the interface. This step was crucial for the application to function properly, as it sets the foundation for further analysis and modeling.

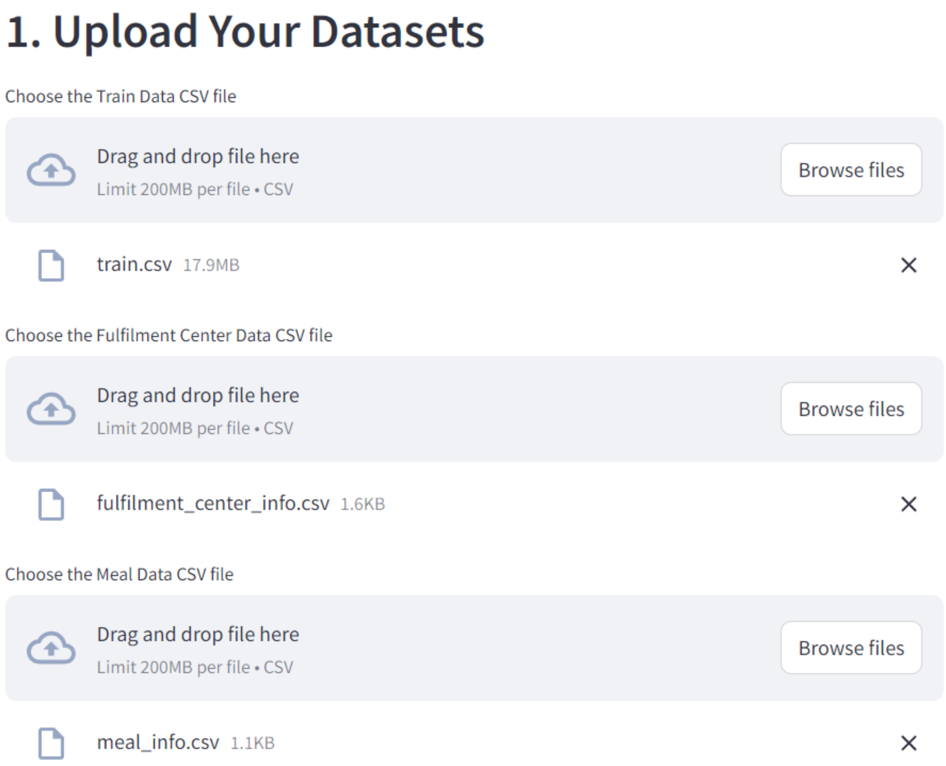

Figure 18: Upload Datasets in Streamlit App

Exploratory Data Analysis

Once the datasets are uploaded, the user can select various EDA operations such as descriptive statistics (e.g., data preview, statistical summary, missing values), time series visualizations, and exploration of relationships between variables through plots like boxplots, histograms, heatmaps, and pairplots. These operations, which were discussed in Chapter 3, provide the user with a comprehensive understanding of the dataset.

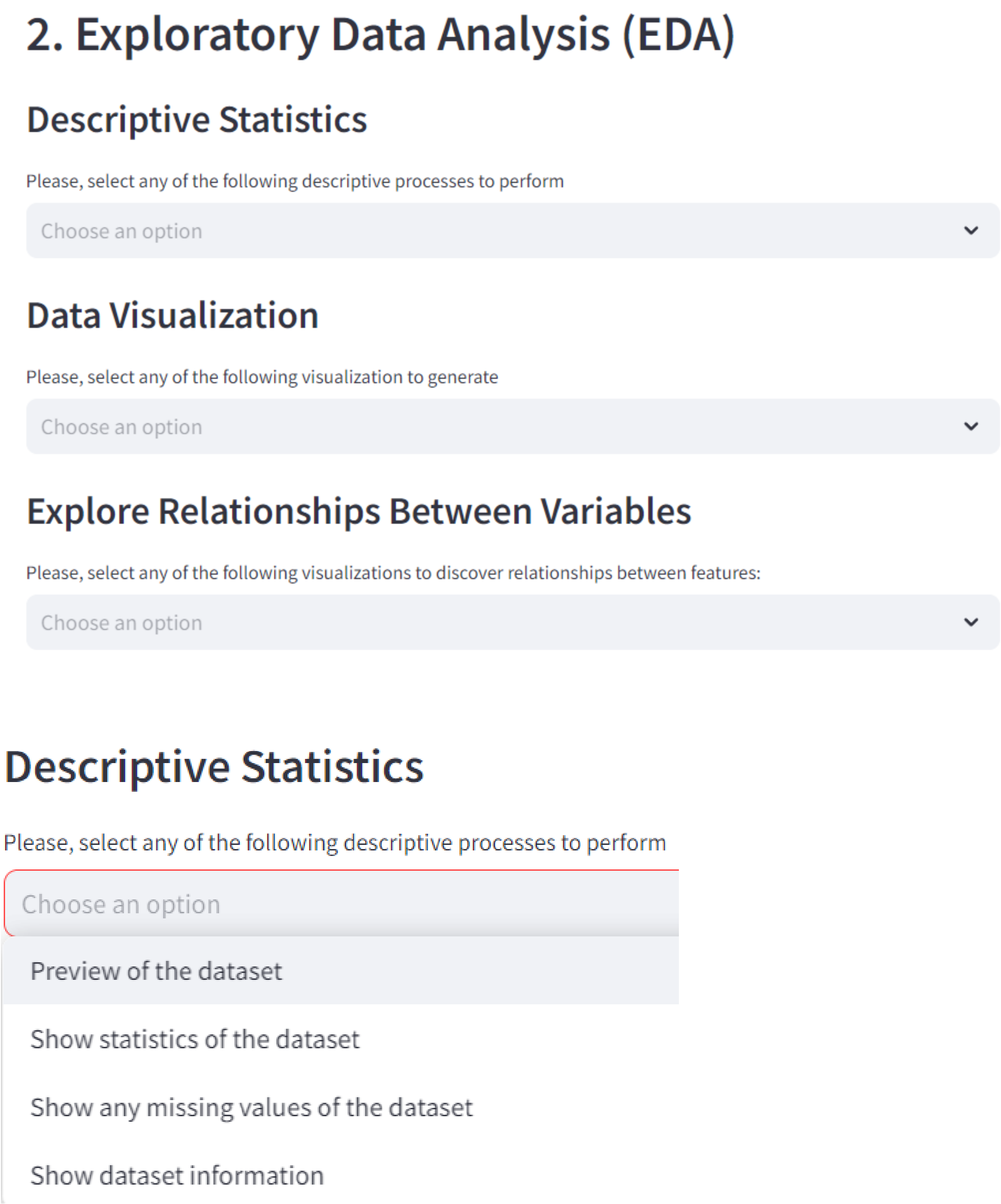

Figure 19: EDA in Streamilt App

Predicting Orders with XGBoost

In this phase, the user can apply the XGBoost model to predict the number of orders for meals over the last 10 days, using historical data. The results of the prediction are displayed clearly to the user.

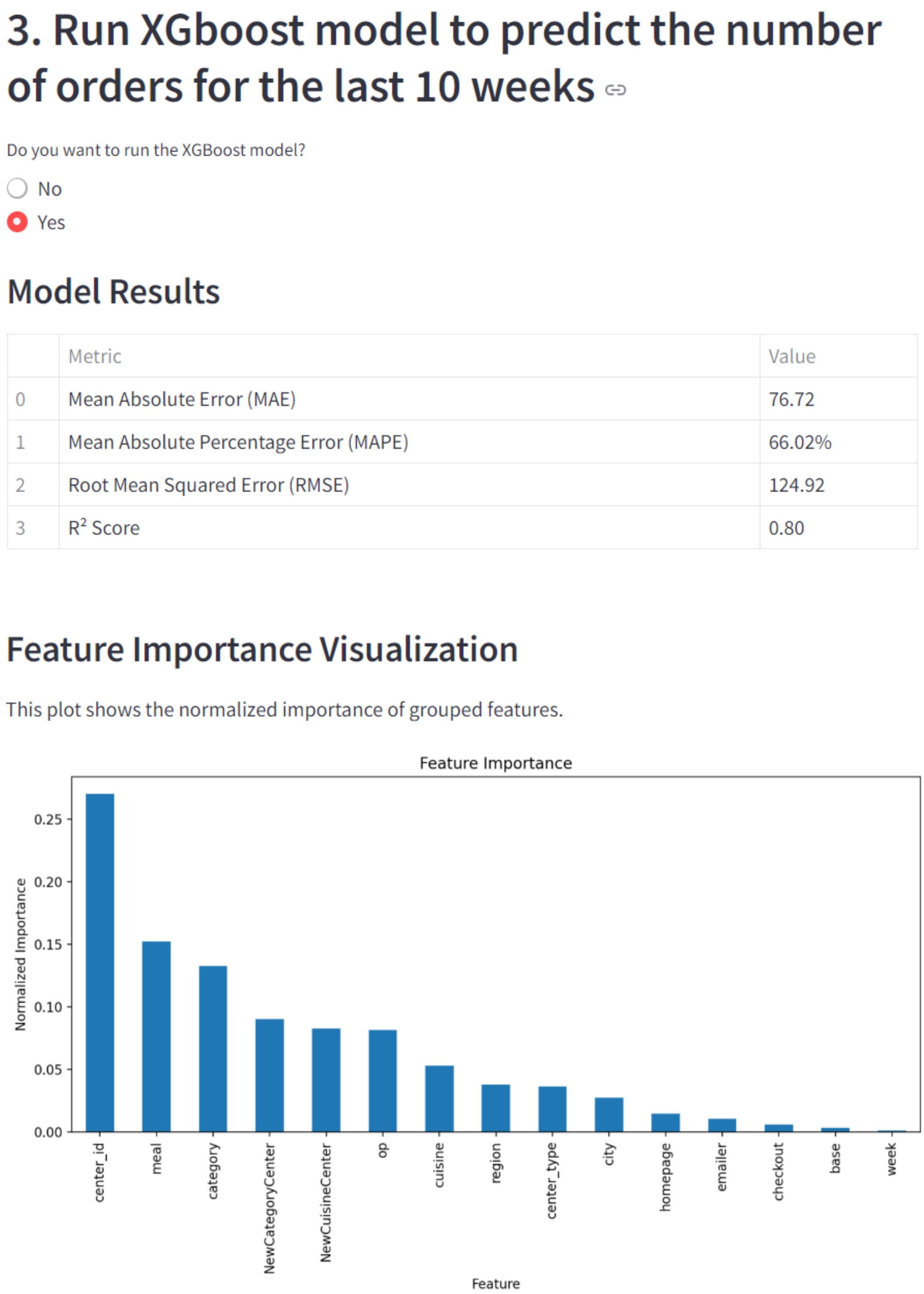

Figure 20: XGBoost Results in Streamlit App

Optimization Model

The final section allows the user to run the optimization model by inputting distribution centers and meals. The optimization algorithm identifies common meals across selected centers and checks if they have predicted demand for the last 10 weeks from week 136 to 145. The user can review the valid meal options and either proceed or adjust their selections.

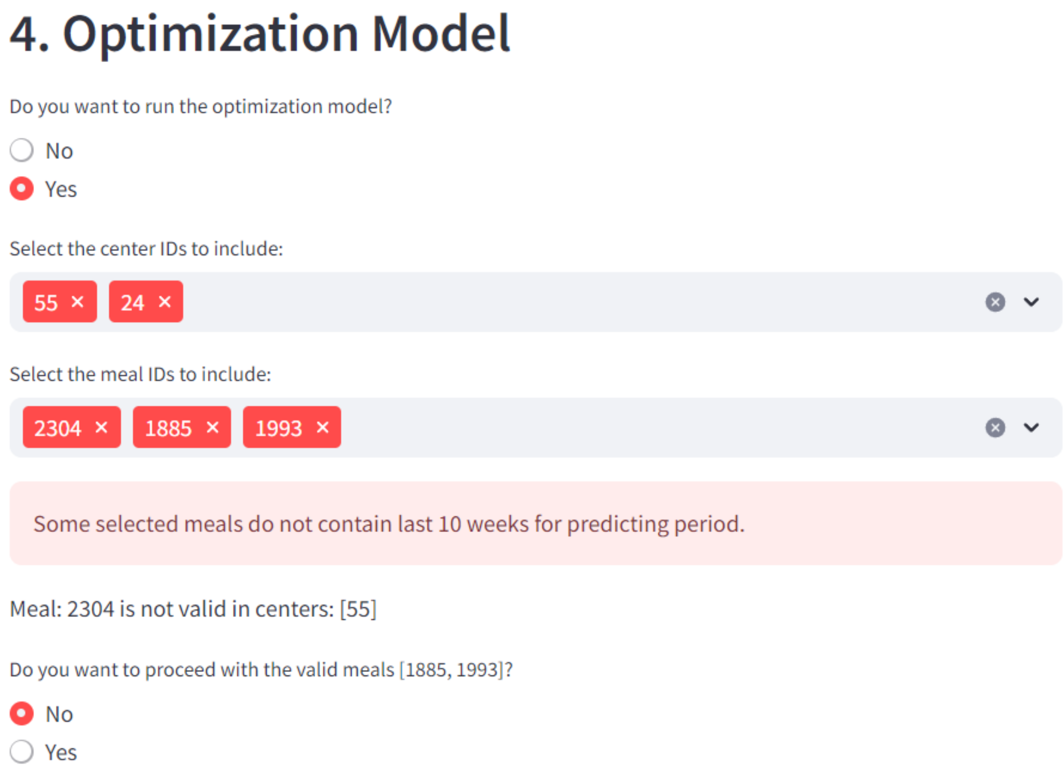

Figure 21: Optimization Model in Streamlit App

Once the user has defined the distribution centers and meals, they enter costs (ordering, holding, unit costs) for each meal. A budget suggestion is provided based on the predicted demand and unit costs, which helps the user understand their budget allocation. The user then inputs specific details, including budget, warehouse capacity, and initial inventory for each meal and center.

Figure 22: Input Costs in Streamlit App

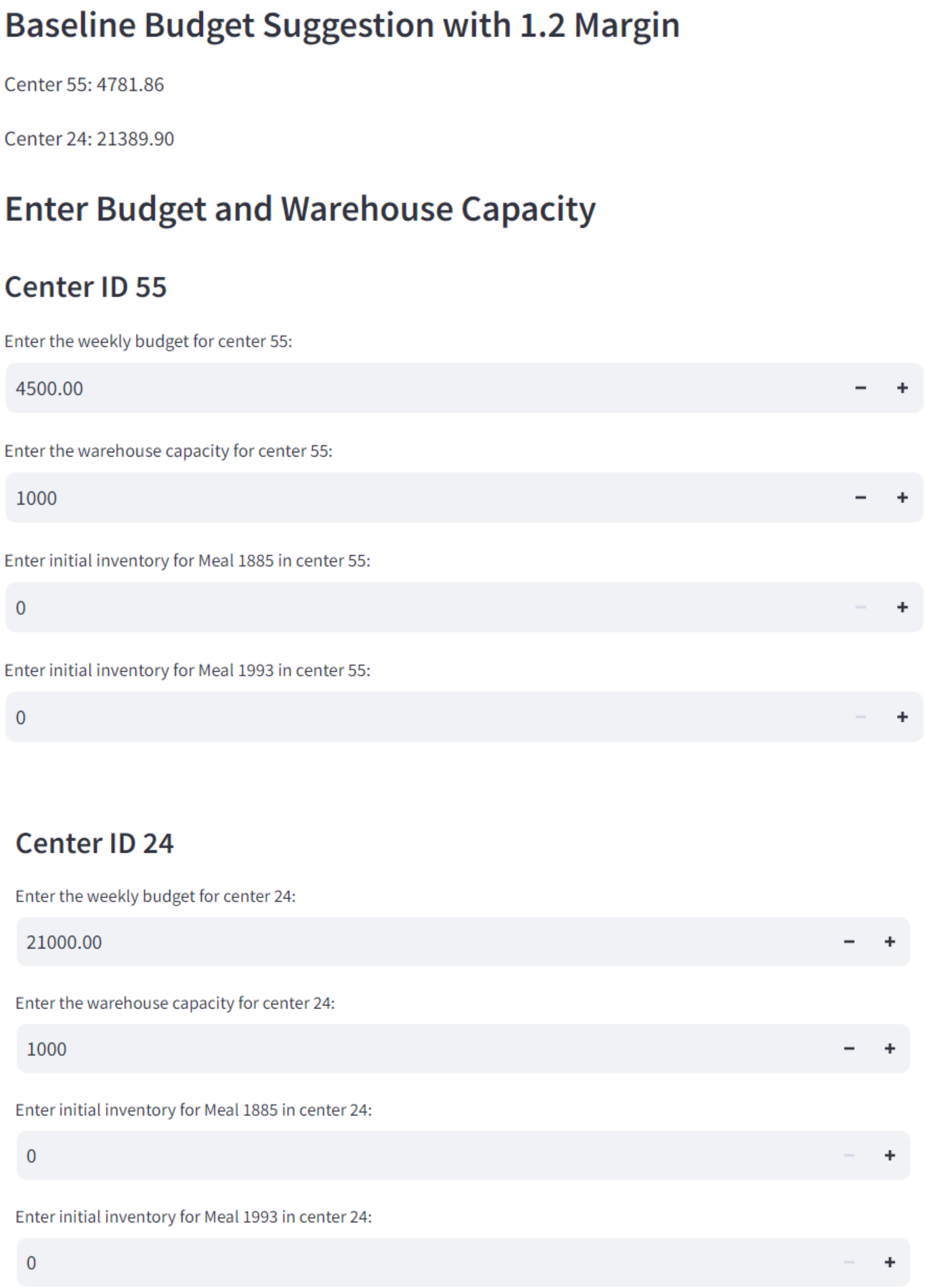

Figure 23: Budgets, Warehouse Capacities and Initial Inventories Inputs in Streamlit

The optimization results are presented through both tables and visualization, providing a comprehensive view of the solution. The results can also be analyzed over time to assess performance and cost-effectiveness.

Figure 24: Optimization Results over Time by Meal and Center

Sensitivity and Simulation Analysis: In addition to optimization, the application offers the ability to perform sensitivity analysis by inputting budget reduction percentages, with results displayed in both tables and plots. Furthermore, the user can simulate different demand variation scenarios and run multiple simulations, viewing the results both in tabular and graphical formats.

Sensitivity Analysis

Do you want to perform a Sensitivity Analysis?

- No
- Yes

Enter budget reduction percentage levels (comma-separated):

0,10,20,30,40,50

Sensitivity Analysis Results:

	Reduction %	Total Cost	Ordering Cost	Holding Cost	Penalty Cost	Total Unmet Demand	Un
0	0	829,926.821	196,294.3852	0	633,632.4358	0	
1	10	1,111,649.0868	196,294.3839	382.2234	914,972.4795	0	
2	20	2,395,759.3583	183,600.0005	0	2,212,159.3579	1,269.4257	
3	30	4,120,688.0382	160,650.0002	0	3,960,038.0381	3,564.4028	
4	40	5,845,616.7274	137,700	0	5,707,916.7274	5,859.3799	
5	50	7,570,545.3801	114,750	0	7,455,795.3801	8,154.3569	

Cost Components and Unmet Demand vs Budget Reduction

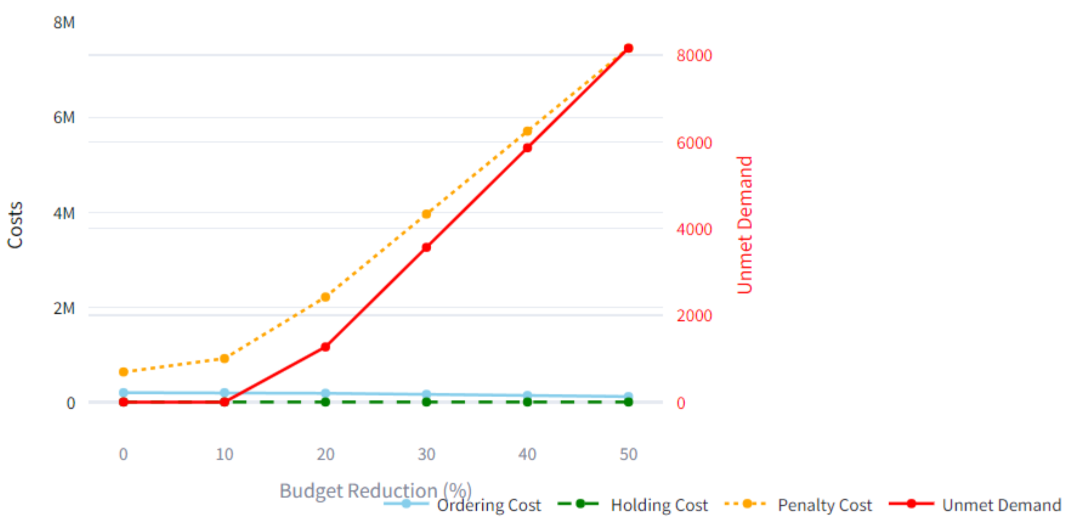

Figure 25: Sensitivity Analysis Results in Streamlit App

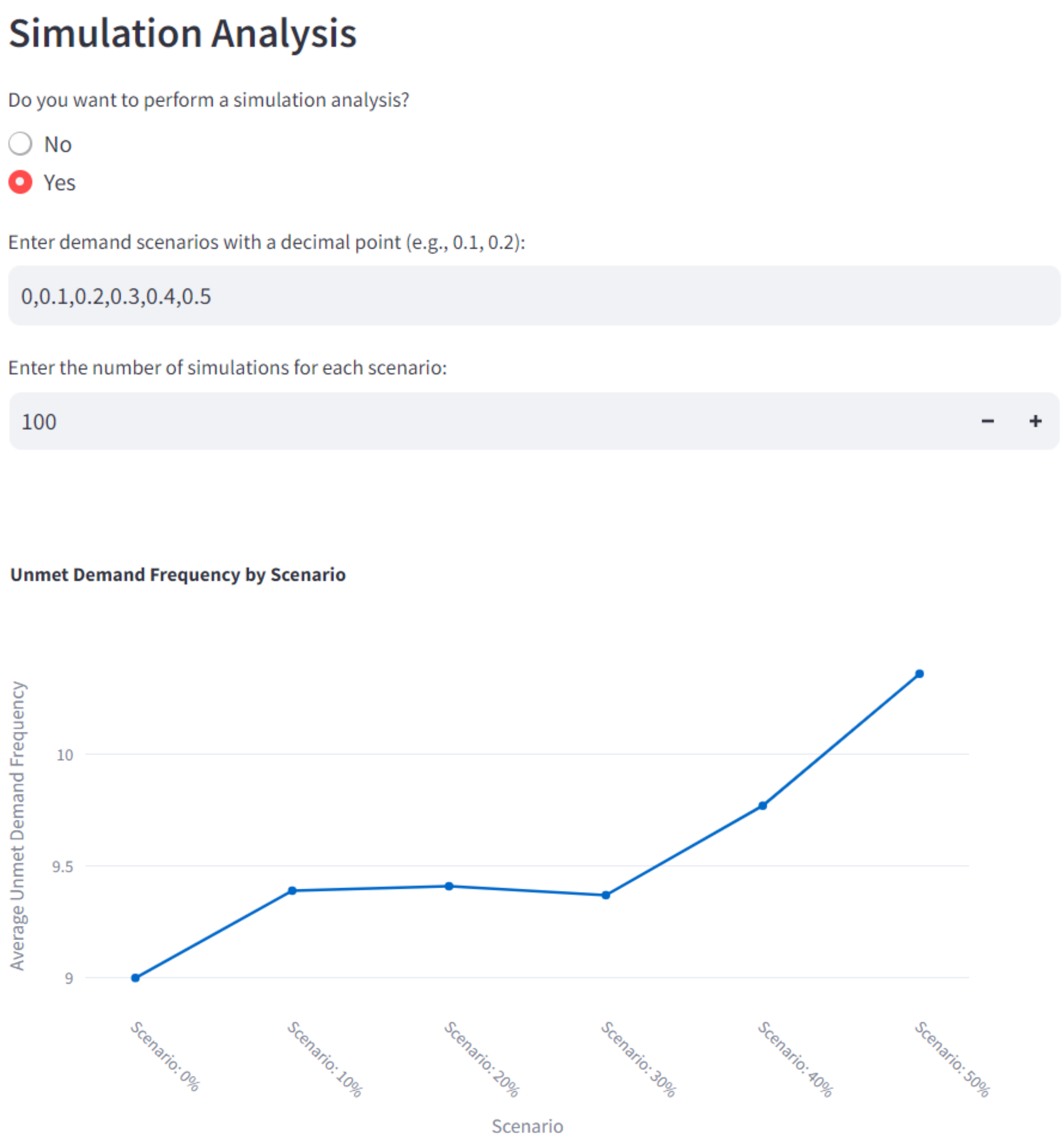

Figure 26: Simulation Analysis Results in Streamlit App

This interactive approach ensures the user has access to a wide range of functionalities, making the web application a powerful tool for predictive modelling, optimization and various insights.

5 Conclusions and Recommendations

Summarization

This study addressed the challenges of crisis-driven food supply chains by developing a web application that integrates advanced forecasting and optimization tools. The XGBoost model demonstrated high accuracy with minimal preprocessing, offering flexibility and easy deployment, making it accessible to non-technical users. The Linear Programming (LP) model adds versatility by accommodating various real-world constraints, while budget sensitivity and demand variation analyses allow users to explore parameter impacts on outcomes. The integration of these models into a dynamic and interactive web application enables real-time decision-making, offering actionable insights to users. The SLR findings also provide valuable insights into state-of-the-art machine learning methods and EOQ models applicable to crisis-driven food supply chains. This way fills a gap existing in the recent literature. In an era of increasing disruptions from pandemics, geopolitical conflicts, and environmental crises, tools like XGBoost and LP models prove essential. The web application developed here serves as a practical, data-driven tool for public and private sector stakeholders, aiding proactive decision-making and mitigating unexpected disruptions in food supply chains. This thesis bridges theoretical advancements with practical applications, benefiting individuals and organizations alike.

Limitations

This thesis utilized Python programming within Google Colab's paid version to access advanced GPU and high-memory RAM resources, ensuring optimal hyperparameter tuning and seamless web application functionality. The free version of Google Colab lacks sufficient computational capacity, leading to crashes during code execution. Similarly, the web application is not publicly accessible, as the free version of Streamlit Community Cloud provides limited GPU and RAM resources, which are inadequate for running the app. Public deployment would require purchasing a paid version or hosting on a cloud platform with high-resource availability. Although the dataset used in this study was large, its authenticity could not be validated, which may affect the generalizability of the findings. The web application can handle large datasets if deployed on high-resource systems, ensuring functionality for diverse and complex data. The Systematic Literature Review (SLR) focused on crisis-driven food supply chains to address a significant gap in the literature. No prior studies were found that jointly explored food demand forecasting and Economic Order Quantity (EOQ) optimization within this specific context. While this focus was essential to address the research gap, it may limit the applicability of the findings to non-crisis scenarios. The literature search spanned three databases. Moreover, the results may reflect a degree of researcher bias using subjective criteria. While the study developed a robust forecasting model (XGBoost), only one machine learning algorithm was explored, despite the potential of other models to deliver comparable or even superior results.

Recommendations

Future research should expand the literature search to include additional databases and maybe relax filtering criteria to uncover more studies addressing machine learning and EOQ methods for crisis-driven food supply chains. Exploring alternative machine learning and deep learning methods, LSTM or Transformers, may yield improved results, provided robust evaluation methods prevent data leakage. Incorporating more real-world constraints into the Linear Programming model, such as transportation variability or dynamic budgets, can enhance its applicability to real-world crises. Validation of the models and web application in real-world settings is crucial to assess their practicality and identify limitations beyond simulations. Additionally, developing scalable, publicly accessible versions of the application and addressing ethical and social considerations can broaden its impact and adoption across diverse stakeholders.

6 Bibliography

[1] S. Çakmakçı, B. Polatoğlu και R. Çakmakçı, «Foods of the Future: Challenges, Opportunities, Trends, and Expectations,» 2024.

[2] S. Çakmakçı, B. Polato και R. Çakmakçı, «Foods of the Future: Challenges, Opportunities, Trends,,» 2024.

[3] INFLUENCING FOOD ENVIRONMENTS FOR HEALTHY DIETS, 2016.

[4] J. J. Rojas-Reyes, L. R. Cadavid και D. L. Pe˜na-Orozco, Disruptions in the food supply chain: A literature review, 2024.

[5] C. Hawkes και M. T. Ruel, Value chains for nutrition, 2011.

[6] Enfoques agroecológicos y otros enfoques innovadores, HLPE INFORME, 2019.

[7] «Nutrition and food systems. A report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security. September 2017. HLPE Report 12,» FAO, 2017.

[8] Z. Irani και A. M. Sharif, «Food security across the enterprise: a puzzle, problem or mess for a circular economy?,» *Journal of Enterprise Information Management*, 2018.

[9] E. Bottani, T. Murino, M. Schiavo και R. Akkerman, «Resilient food supply chain design: Modelling framework and metaheuristic solution approach,» 2019.

[10] V. R. REDDY, S. K. SINGH και V. ANBUMOZHI, «Food Supply Chain Disruption due to Natural Disasters: Entities, Risks, and Strategies for Resilience,» 2016.

[11] M. Najafi, K. Eshghi και W. Dullaer, «A multi-objective robust optimization model for logistics planning in the earthquake response phase,» 2013.

[12] F. Rayat, M. Musavi και A. Bozorgi-Amiri, «Bi-objective reliable location-inventory-routing problem with partial backordering under disruption risks: A modified AMOSA approach,» 2017.

[13] Y. Zhang, M. Qi, W.-H. Lin και L. Miao, «A metaheuristic approach to the reliable location routing problem under disruptions,» 2015.

[14] K. Kraemer, J. Cordaro, J. Fanzo, M. Gibney, E. Kennedy, A. Labrique, J. Steffen και M. Eggersdorf, «45 - 59: Chapter 1.3 Managing Value Chains for Improved Nutrition,» 2016.

[15] «Agricultural trade & policy responses during the first wave of the COVID-19 pandemic in 2020,» 2021.

[16] A. A Hecht, E. Biehl, D. J Barnett και R. A Neff, «Urban Food Supply Chain Resilience for Crises Threatening Food Security: A Qualitative Study,» 2018.

[17] C. Iacovacci, M. Caratù, G. Addamo και V. Grasso, «Analyzing the Impact of COVID-19 on Business Performance through the Case-Study of a Green Italian Start-Up,» 2023.

[18] G. Romagnoli, M. Gallo, A. Liccardo και R. Riedel, «Guest editorial: Opportunities and threats in providing remote access to manufacturing-related environments,» 2023.

[19] N. A. Mostafa, A. A. Hussein, M. Elsheeta και G. Romagnoli, «Impacts of COVID-19 and the Russian–Ukrainian Conflict on Food Supply Chain: A Case Study from Bread Supply Chain in Egypt,» 2024.

[20] A. P. Found, D. Mogale και Z. Xu, «Food supply chain resilience in major disruptions,» *Journal of Manufacturing Technology Management,* 2024.

[21] H. H. WANG και H. Na, «Panic buying? Food hoarding during the pandemic period with city lockdown,» *Journal of Integrative Agriculture,* 2020.

[22] H. T. Y. Yoshizaki, I. d. B. Juni, C. M. Hino, L. L. Aguiar και M. C. Rodri, «Relationship between Panic Buying and Per Capita Income during COVID-19,» 2020.

[23] M. Hall, P. Fieger, G. Prayag και D. Dyas, «Panic Buying and Consumption Displacement during COVID-19: Evidence from New Zealand,» 2021.

[24] L. D. Montenegro και M. N. Young, «Operational Challenges in the Food Industry and Supply Chain during the COVID-19 Pandemic: A Literature Review,» 2020.

[25] I. Brito Jr, H. Yoshizaki, F. Azevedo, N. d. C. Bruno, R. Fray da Silva, C. Hino, L. L. Aguiar και I. M. Ferreira de Ataide, «Panic Buying Behavior Analysis according to Consumer Income and Product Type during COVID-19,» 2023.

[26] C. P. Boyacı-Gündüz, S. A. Ibrahim, O. Chien Wei και C. M. Galanakis, «Transformation of the Food Sector: Security and Resilience during the COVID-19 Pandemic,» 2021.

[27] J. E. Hobbs, «Food supply chain resilience and the COVID-19 pandemic: What have we learned?,» 2021.

[28] M. E. Maqboola, . A. Farhanb και M. A. Qamar, «Global impact of COVID-19 on food safety and environmental sustainability: Pathways to face the pandemic crisis,» 2024.

[29] F. Jawed, S. Manazir και A. Zehra, «The novel Coronavirus disease (COVID-19) pandemic: Knowledge, attitude, practice, and perceived stress among health care workers in Karachi, Pakistan,» 2020.

[30] M. Rizou, I. M Galanakis, T. MS Aldawoud και C. M Galanakis, «Safety of foods, food supply chain and environment within the COVID-19 pandemic,» 2020.

[31] P. Sharma, H. Kaur, M. Sharma και V. Sahore , «A review on applicability of naturally available adsorbents for the removal of hazardous dyes from aqueous waste,» 2011.

[32] J. Doward, «Fruit and veg ‘will run out’ unless Britain charters planes to fly in farm workers from eastern Europe,» 2020.

[33] «COVID-19 Analytical Snapshots».

[34] N. A. Mostafa, A. A. Hussein, M. Elsheeta και G. Romagnoli, «Impacts of COVID-19 and the Russian–Ukrainian Conflict on,» 2024.

[35] H. M. Kabir, R. M. Rabbani και A. Yomna, «Socioeconomic Impact of COVID-19 in MENA region and the Role of Islamic Finance,» 2021.

[36] R. S. Gray, «Agriculture, transportation, and the COVID-19 crisis,» 2020.

[37] Z. Nakat και C. Bou-Mitri, «COVID-19 and the food industry: Readiness assessment,» 2021.

[38] M. Pu και Y. Zhong, «Rising concerns over agricultural production as COVID-19 spreads: Lessons from China,» 2020.

[39] C. Fontan Sers και M. Mughal, «Covid-19 outbreak and the need for rice self-sufficiency in West Africa,» 2020.

[40] S. Bakalis, V. P. Valdramidis, D. Argyropoulos, L. Ahrne, J. Chen, P. Cullen, E. Cummins, A. K. Datta, C. Emmanouilidis, T. Foster, P. J. Fryer, O. Gouseti, A. Hospido, K. Knoerzer, A. LeBail, A. G. Marangoni, P. Rao, O. K. Schlüter, P. Taoukis, E. Xanthakis και J. F. V. Impe, «Perspectives from CO+RE: How COVID-19 changed our food systems and food security paradigms,» 2020.

[41] L. M Hill, H. Davis, M. Drewry, V. Shelus, S. M Bartels, K. Gora Combs, K. M Ribis και A. J Lazard, «Barriers to and Facilitators of COVID-19 Prevention Behaviors Among North Carolina Residents,» 2022.

[42] H. Onyeaka, . U. M. Nwauzoma και A. A. Akinsem, «The ripple effects of climate change on agricultural sustainability and food security in Africa,» 2024.

[43] « NOAA National Centers for Environmental Information. Monthly Global Climate Report for Annual 2023,» 2023.

[44] X. Yuan, S. Li, C. Jinliang, H. Yu, T. Yang, C. Wang, S. Huang, C. Haochong και X. Ao, «Impacts of Global Climate Change on Agricultural Production: A Comprehensive Review,» 2024.

[45] V. Owino, C. Kumwenda, B. Ekesa, M. E. Parker, L. Ewoldt, N. Roos, W. T. Lee και D. Tome, «The impact of climate change on food systems, diet quality, nutrition, and health outcomes: A narrative review,» 2022.

[46] R. Subhashni, S. Roodbar, B. Catherine και D. W. Wolfe, «Food Security and Climate Change: Differences in Impacts and Adaptation Strategies for Rural Communities in the Global South and North,» 2022.

[47] P. Aggarwal, P. Joshi, J. Ingram και R. Gupta, «Adapting food systems of the Indo-Gangetic plains to global environmental change: key information needs to improve policy formulation,» 2004.

[48] M. Jaafar, F. B. Abdelaziz και B. Maddah, «Effect of COVID-19 on Supply Chain Decisions,» 2021.

[49] K. F. Davis, S. Downs και J. A. Geph, «Towards food supply chain resilience to environmental shocks,» 2020.

[50] A. S. A. k. Kiani, Y. Kuslu και A. BILGIC, «Examining the Role of Livelihood Diversification as a Part of Climate-Smart Agriculture (CSA) Strategy (İklim-Akıllı Tarım stratejisinin bir parçası olarak geçim çeşitliliğinin rolünün incelenmesi),» 2020.

[51] U. Sarker, S. Oba, R. Ullah, A. Bari, S. Ercisli, S. Skrovankova, A. Adamkova, M. Zvonkova και J. Mlcek, «Nutritional and bioactive properties and antioxidant potential of Amaranthus tricolor, A. lividus, A viridis, and A. spinosus leafy vegetables,» 2024.

[52] T. Hasegawa, G. Sakurai, S. Fujimori, K. Takahashi, Y. Hijioka και T. Masui, «Extreme climate events increase risk of global food insecurity and adaptation needs,» 2021.

[53] A. S. López, P. Garcia-Perez, M. C. Rodríguez, P. G. Oliveira, P. Otero, M. F. Corral, H. Cao, M. A. Prieto και J. Simal-Gandara, «Challenges for future food systems: From the Green Revolution to food supply chains with a special focus on sustainability,» 2022.

[54] N. Marmai, M. F. Villoria και M. Guerzoni, «How the Black Swan damages the harvest: Extreme weather events and the fragility of agriculture in development countries,» 2022.

[55] S. B. ILUGBUSI, R. A. B, A. A. John και A. OGUNDELE, «Financial Liberalization and Economic Growth in Nigeria (1986-2018),» 2020.

[56] O. I. Oriekho, O. Adisa και S. Ilugbusi, «CLIMATE CHANGE AND FOOD SUPPLY CHAIN ECONOMICS: A COMPREHENSIVE ANALYSIS OF IMPACTS, ADAPTATIONS, AND SUSTAINABILITY,» 2024.

[57] P. Haessner, J. Haessner και M. McMurtrey, «Trends & Challenges in the Food Supply Chain,» 2024.

[58] B. Małgorzata και B. Hanna, «Direct and Indirect Consequences of the War in Ukraine for Polish Trade in Agri-Food Products,» 2023.

[59] J. Sandeep, H. Trollman, F. Trollman, G. Garcia-Garcia, C. Parra-López, L. Duong, W. Martindale, P. E. S. Munekata, J. M. Lorenzo, A. Hdaifeh, A. Hassoun, K. Salonitis και M. Afy-Shararah, «The Russia-Ukraine Conflict: Its Implications for the Global Food Supply Chains,» 2022.

[60] J. Horton και D. Palumbo, «Russia sanctions: What impact have they had on its oil and gas exports?,» 2023.

[61] «Global food losses and food waste,» 2011.

[62] M. Rodrigues, V. Miguéis, S. Freitas και T. Machado, «Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste,» *Journal of Cleaner Production,* 2024.

[63] F. Girotto, L. Alibardi και R. Cossu, «Food waste generation and industrial uses: A review,» 2015.

[64] S. S. Kamble, A. Gunasekaran και S. A. Gawankar, «Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications,» *International Journal of Production Economics,* 2020.

[65] C. Martin-Rios, C. Demen-Meier, S. Gössling και C. Cornuz, «Food waste management innovations in the foodservice industry,» *Waste Management,* 2018.

[66] C. Priefer, J. Jörissen και K.-R. Bräutigam, «Food waste prevention in Europe – A cause-driven approach to identify the most relevant leverage points for action,» *Resources, Conservation and Recycling,* 2016.

[67] R. Messner, C. Richards και H. Johnson, «The "Prevention Paradox": food waste prevention and the quandary of systemic surplus production,» 2020.

[68] X. Zhang, J. Bao, S. Xu, Y. Wang και S. Wang, «Prediction of China's Grain Consumption from the Perspective of Sustainable Development—Based on GM(1,1) Model,» 2022.

[69] E. Fukase και W. Martin, «Who Will Feed China in the 21st Century? Income Growth and Food Demand and Supply in China,» 2015.

[70] J.-k. HUANG, W. WEI, Q. CUI και W. XIE, «The prospects for China's food security and imports: Will China starve the world via imports?,» 2017.

[71] S. Fan, W. Si και Y. Zhang, «How to prevent a global food and nutrition security crisis under COVID-19?,» 2020.

[72] C. P. Yale, H. T. Y. Yoshiza και F. L. Paulo, «A New Zero-Inflated Negative Binomial Multilevel Model for Forecasting the Demand of Disaster Relief Supplies in the State of Sao Paulo, Brazil,» 2022.

[73] «The State of Food Insecurity in the World 2009 Economic crices - impacts and lessons learned,» Rome, 2009.

[74] F. I. U. W. και W. , The State of Food Security and Nutrition in the World 2023 Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum, Rome, Italy, 2023.

[75] A. Duisenbekova, M. Kulisz, A. Danilowska, A. Gola και M. Ryspekova, «Predicting Food Consumption to Reduce the Risk of Food Insecurity in Kazakhstan,» 2023.

[76] A. Coleman-Jensen, M. P. Rabbitt, C. A. Gregory και A. Singh, «Household Food Security in the United States in 2021,» 2022.

[77] M. F. Rabbi, M. Hasan και S. Kovács, «Food Security and Transition towards Sustainability,» 2021.

[78] B. AIMURZINA, M. KAMENOVA, A. OMAROVA, G. PESTUNOVA, A. KARIPOVA και K. MADENOVA, «Particular Forms of Management of Agro – Industrial Complex in the Sustainable Development of Agriculture,» 2018.

[79] M. ACI και D. YERGÖK, «Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory,» 2023.

[80] S. Punia και S. Shankar, «Predictive analytics for demand forecasting: A deep learning-based decision support system,» 2022.

[81] T. Wang, «A K-means Group Division and LSTM Based Method for Hotel Demand Forecasting,» 2021.

[82] M. Karatas, «Reducing Demand Signal Variability via a Quantitative Fuzzy Grey Regression Approach,» 2018.

[83] K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk και A. Kozina, «Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network,» 2021.

[84] «UNEP Food Waste Index Report 2021,» 2021.

[85] M. Meneghini, M. Anzanello, A. Kahmann και G. Tortorella, «Quantitative demand forecasting adjustment based on qualitative factors: case study at a fast food restaurant,» 2018.

[86] W. v. d. Aalst , «Data Science in Action,» 2016.

[87] S. K. Panda και S. N. Mohanty, «Time Series Forecasting and Modeling of Food Demand Supply Chain Based on Regressors Analysis,» 2023.

[88] M. R. Adzaky, R. Erlina και D. A. Siti Ambarw, «Analysis of Raw Material Inventory Control with Using Economic Order Quantity (EOQ) Method,» 2024.

[89] A. W. Yuniasih και N. R. Lufti A'yuni, «Literature Review of Inventory with Probabilistic Economic Order Quantity (EOQ),» 2024.

[90] S. Senthilnathan, «Economic Order Quantity (EOQ),» 2019.

[91] Y. Xiao και M. Watson, «Guidance on Conducting a Systematic Literature Review,» 2017.

[92] A. Boland, G. Cherry και R. Dickson, «Doing a Systematic Review: A Student′s Guide,» 2017.

[93] J. Webster και R. T. Watson, «Analyzing the Past to Prepare for the Future: Writing a Literature Review,» 2002.

[94] Y. Levy και T. J. Ellis, «A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research,» 2006.

[95] G. Paré, M.-C. Trudel, M. Jaana και S. Kitsiou, «Synthesizing information systems knowledge: A typology of literature reviews,» 2015.

[96] E. M. Antman, J. Lau, B. Kupelnick, F. Mosteller και T. C. Chalmers, «A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction,» 1992.

[97] A. D. Oxman και G. H. Guyatt, «The science of reviewing research,» 1993.

[98] A. Penedones, C. Alves και F. Batel-Marques, «Recommendations to conduct and reportsystematic reviews in medical literature: ascoping review,» 2019.

[99] C. Cronin, «Doing your literature review: traditional and systematic techniques,» 2011.

[100] M. Templier και G. Paré, «A Framework for Guiding and Evaluating Literature Reviews,» 2015.

[101] D. M. Rousseau, J. Manning και D. Denyer, «11 Evidence in Management and Organizational Science: Assembling the Field's Full Weight of Scientific Knowledge Through Syntheses,» 2007.

[102] J. Thomas και A. Harden, «Methods for the thematic synthesis of qualitative research in systematic reviews,» 2008.

[103] P. Brereton, B. Kitchenham, D. Budgen, M. Turner και M. Khalil, «Lessons from applying the systematic literature review process within the software engineering domain. J Syst Softw,» 2007.

[104] B. Kitchenham και S. M. Charters, «Guidelines for performing Systematic Literature Reviews in Software Engineering,» 2006.

[105] «Scopus,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Scopus.

[106] «ProQuest,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/ProQuest.

[107] «IEEE,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Institute_of_Electrical_and_Electronics_Engineers.

[108] C. Okoli και K. Schabram, «A Guide to Conducting a Systematic Literature Review of Information Systems Research,» 2010.

[109] M. Peters, C. M. Godfrey, H. Khalil, P. Mcinerney, D. Parker και C. B. Soares, «Guidance for conducting systematic scoping reviews,» 2015.

[110] M. Noordzij, L. Hooft, F. W. Dekker, C. Zoccali και K. J. Jager, «Systematic reviews and meta-analyses: when they are useful and when to be careful,» 2009.

[111] J. Rowley και F. Slack, «Conducting a literature review,» 2004.

[112] R. Whittemore και K. Knafl, «The integrative review: Update methodology,» 2005.

[113] E. Anitha, M. N. Prasath, L. Sanjai, J. A. Andrea Shiny και P. Varsini, «Effective Food Demand Forecasting Using Machine Learning Algorithms,» 2023.

[114] M. Rodrigues, V. Miguéis, S. Freitas και T. Machado, «Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste,» 2024.

[115] C. Zhu, J. Wang και J. Tian, «Grain Consumption Forecasting: One Modified MLR Model Combined with Time Series Forecasting Theory,» 2020.

[116] S. Paul και L. B. Davis, «An ensemble forecasting model for predicting contribution of food donors based on supply behavior,» 2022.

[117] U. Bozdoğan και G. I. Alptekin, «Demand Forecasting for Daily Retail Orders in Fresh Food Market,» 2023.

[118] N. K. Pandey, A. K. Mishra, V. Kumar, A. Kumar, M. Diwakar και N. Tripathi, «Machine Learning based Food Demand Estimation for Restaurants,» 2023.

[119] S. Chatterjee, «A Stack-based Ensemble Model with Explainability for Food Demand Forecasting,» 2022.

[120] J. Zhao, M. Li, J. Zhang και F. Qi, «A Demand Forecasting Method for High Value-Added Agri - Food Based on Machine Learning and Time Series Analysis,» 2023.

[121] L. Neubauer και P. Filzmoser, «Improving forecasts for heterogeneous time series by "averaging", with application to food demand forecasts,» 2024.

[122] M. K. H. Mondal, A. Mondal, S. Chakraborty, K. Shubhranshu, A. K. Jha και M. K. Roy, «Advanced Deep Learning and NLP for Enhanced Food Delivery: Future Insights on Demand Prediction, Route Optimization, Personalization, and Customer Support,» 2023.

[123] V. Kuppulakshmi, C. Sugapriya και N. D. Pillai, «An inventory model for fish marketing under uncertain lockdown situation and normal backordering situation,» 2022.

[124] A. T. Rahman και D. Widyaningrum, «Analysis of Inventory Control of Perishable Goods with Capital Constraints and Warehouse Capacity Using the Lagrange EOQ Method (Case Study: UD. XYZ),» 2023.

[125] R. M. Sari, S. Y. Firiza, Erwin, K. Syahputri, I. Rizkya και I. Siregar, «Minimizing inventory cost of dried food materials availability: An analyzing in teaching hospital,» 2021.

[126] H. Meileni, D. J. Putra, D. Apriyanti, I. Satriadi και S. Oktapriandi, «Inventory Of Goods Data Processing Using The Economic Order Quantity (Eoq) Method,» 2020.

[127] C. Çalı¸skan, «A Comparison of Simple Closed-Form Solutions for the EOQ Problem for Exponentially Deteriorating Items,» 2022.

[128] A. Brijith, «Data Preprocessing for Machine Learning,» 2023.

[129] Available: https://medium.com/@sunnykumar1516/what-is-label-encoding-application-of-label-encoder-in-machine-learning-and-deep-learning-models-c593669483ed.

[130] Available: https://contrib.scikit-learn.org/category_encoders/binary.html#category_encoders.binary.BinaryEncoder.

[131] Available: https://medium.com/@kevin.menear/comparing-label-encoding-one-hot-encoding-and-binary-encoding-for-handling-categorical-variables-933544ccbd02.

[132] Available: https://contrib.scikit-learn.org/category_encoders/onehot.html#category_encoders.one_hot.OneHotEncoder.

[133] Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html.

[134] Available: https://xgboost.readthedocs.io/en/latest/python/python_api.html.

[135] Sachinsoni, 2023. Available: https://medium.com/@sachinsoni600517/model-evaluation-techniques-in-machine-learning-47ae9fb0ad33.

[136] Available: https://scikit-learn.org/dev/modules/generated/sklearn.model_selection.TimeSeriesSplit.html.

[137] Available: https://scikit-learn.org/1.5/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV.

[138] Available: https://scikit-learn.org/1.5/modules/model_evaluation.html#r2-score-the-coefficient-of-determination.

[139] Available: https://vitalflux.com/mse-vs-rmse-vs-mae-vs-mape-vs-r-squared-when-to-use/.

[140] «Medium [GeorgeKamtziridis],» Available: https://medium.com/@geokam/time-series-forecasting-with-xgboost-and-lightgbm-predicting-energy-consumption-with-lag-features-dbf69970a90f.

[141] 2023. Available: https://www.geeksforgeeks.org/xgboost/.

[142] T. Chen και C. Guestrin, «XGBoost: A Scalable Tree Boosting System,» 2016.

[143] Available: https://www.nvidia.com/en-eu/glossary/xgboost/.

[144] E. Kavlakoglu και E. Russi, 2024. Available: https://www.ibm.com/topics/xgboost.

[145] Available: https://xgboost.readthedocs.io/en/latest/index.html.

[146] A. C. Sánchez, «Comparison of Pearson and Spearman Correlation for didactic purposes,» 2023.

[147] Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html.

[148] Available: https://coin-or.github.io/pulp/.

[149] Available: https://en.wikipedia.org/wiki/Sensitivity_analysis.

[150] E. Castillo, R. Mínguez και C. Castillo, «Sensitivity analysis in optimization and reliability problems,» 2008.

[151] Available: https://www.twi-global.com/technical-knowledge/faqs/faq-what-is-simulation#WhatDoesitMean.

[152] Available: https://en.wikipedia.org/wiki/Simulation.

[153] Available: https://streamlit.io/.

[154] Available: https://streamlit.io/cloud.

[155] Available: https://ngrok.com/.