DESIGN AND PERFORMANCE EVALUATION OF DEMAND FORECASTING SYSTEM FOR ONLINE FOOD DATA

(SANAL GIDA VERİSİ ÜZERİNDE TALEP TAHMİN SİSTEMİ TASARIMI VE BAŞARIM DEĞERLENDİRMESİ)

by

Meltem ARSLAN, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

July 2024

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have supported and guided me throughout the process of completing this thesis.

First and foremost, I would like to thank my supervisor, Gülfem ALPTEKİN, for their invaluable guidance, continuous support, and patience. Their insightful feedback and encouragement have been essential in the successful completion of this work.

I am deeply thankful to my family for their unwavering support and encouragement. Their belief in me has been a constant source of motivation throughout my academic journey.

Thank you all for your contributions, encouragement, and unwavering support. This work would not have been possible without you.

July 2024
Meltem ARSLAN

Table of Contents

LIST OF SYMBOLS

EU : European Union
SDG : Sustainable Development Goals
TÜBİTAK : The Scientific and Technological Research Council of Turkey
RFR : Random Forest Regressor
GBR : Gradient Boosting Regressor
LightGBM : Light Gradient Boosting Machine Regressor
XGBoost : Extreme Gradient Boosting Regressor
LSTM : Long Short-Term Memory
NARX : Nonlinear Autoregressive with Exogenous Inputs
NARXNN : Nonlinear autoregressive exogenous neural network
RNN : Recurrent Neural Network
MAE : Mean Absolute Error
RMSE : Root Mean Squared Error
MSE : Mean Squared Error
ARIMA : AutoRegressive Integrated Moving Average
ANN : Artificial Neural Network
CNN : Convolutional Neural Network
SVR : Support Vector Regression
LR : Linear Regression
DTR : Decision Tree Regressor
RFR : Random Forest Regressor
MLP : Multi-layer Perceptron
GOSS : Gradient-based One-Side Sampling
EFB : Exclusive Feature Bundling
R^2 : R-square

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Demand prediction is a critical component in optimizing supply chain management and enhancing operational efficiency. By predicting demand, manufacturers can plan their production, make informed decisions regarding distribution and storage options. However, demand prediction in the fresh food industry presents significant challenges due to the short shelf life and high demand variability. Additionally, external factors like weather conditions and seasonal changes make predicting demand for fresh food even more complex. These complexities necessitate sophisticated prediction models that can adapt to rapid changes and incorporate real-time data. Implementing such models can help mitigate waste, optimize inventory levels, and ultimately improve customer satisfaction. This thesis explores advanced methodologies and algorithms seven for accurately predicting demand, particularly focusing on online food market data. For this purpose, we used two food related datasets; Online Market Dataset and Food Demand Forecasting Dataset. As models we used Linear Regression, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, XGBoost, Multi-layer Perceptron Regressor, Light Gradient Boosting Machine, and NARX Model. The first seven models were used as a benchmark for the time series model, NARX. The performance metrics used are Root Mean Squared Error, R-squared, Mean Squared Error, and Mean Absolute Error. Among the seven models, for the Online Market Dataset, the Decision Tree Regressor and Random Forest Regressor exhibit the best performance. In addition, for the Food Demand Forecasting dataset, the Light Gradient Boosting Machine model exhibits the best performance. Lastly, compared to other models, the Nonlinear Autoregressive Exogenous Model shows moderate performance.

Keywords: Demand Prediction, NARX

ÖZET

Talep tahmini, tedarik zinciri yönetimini optimize etmede ve operasyonel verimliliği artırmada kritik bir bileşendir. Talebi önceden tahmin ederek, üreticiler üretimlerini planlayabilir, dağıtım ve depolama seçenekleri konusunda bilinçli kararlar alabilir. Ancak, taze gıda endüstrisinde talep tahmini, kısa raf ömrü ve yüksek talep değişkenliği nedeniyle önemli zorluklar sunmaktadır. Ayrıca, hava koşulları ve mevsimsel değişiklikler gibi dış faktörler, taze gıda talebini tahmin etmeyi daha da karmaşık hale getirir. Bu karmaşıklıklar, hızlı değişimlere uyum sağlayabilen ve gerçek zamanlı verileri içerebilen sofistike tahmin modellerini gerektirir. Bu tür modellerin uygulanması, israfı azaltmaya, envanter seviyelerini optimize etmeye ve nihayetinde taze ürünlerin sürekli olarak mevcut olmasını sağlayarak müşteri memnuniyetini artırmaya yardımcı olabilir. Bu tez, özellikle sanal gıda piyasası verilerine odaklanarak talebi doğru bir şekilde tahmin etmek için gelişmiş metodolojileri ve algoritmaları araştırmaktadır. Bu amaçla, bir Türk firması tarafından geliştirilen bir veri seti ve Kaggle'dan Gıda Talebi Tahmini adlı bir veri seti kullandık. Model olarak, Doğrusal Regresyon, Karar Ağacı Regressörü, Rastgele Orman Regressörü, Gradient Boosting Regressörü, XGBoost, Çok Katmanlı Algılayıcı Regressörü, Light Gradient Boosting Makinesi ve NARX Modeli kullandık. İlk yedi model, zaman serisi modeli olan NARX Model için bir kıyaslama olarak kullanıldı. Modellerin performansını değerlendirmek için Kök Ortalama Kare Hatası, R-kare, Ortalama Kare Hatası ve Ortalama Mutlak Hata'dır. Yedi makine öğrenimi modeli arasında, Sanal Market Veri Seti için Karar Ağacı Regressörü ve Rastgele Orman Regressörü en iyi performansı sergilemektedir. Ayrıca, Gıda Talebi Tahmini veri seti için, Light Gradient Boosting Makinesi modeli en iyi performansı sergilemiştir. Son olarak, diğer modellere kıyasla, NARX orta düzeyde performans göstermektedir.

Anahtar Kelimeler: Talep Tahmini, NARX

1. INTRODUCTION

Reducing food waste is a critical component of the European Green Deal, particularly under the "Farm to Fork Strategy," which aims to create a sustainable, fair, healthy and environmentally friendly food system (Farm to Fork Strategy, 2020). The strategy outlines several significant measures to combat food waste. First, the European Commission plans to propose legally binding targets by the end of 2023 to reduce food waste across the European Union (EU). These targets will be informed by the first comprehensive EU-wide monitoring of food waste levels (EU Actions Against Food Waste, n.d.).

Furthermore, the strategy includes revising the EU's date marking rules, such as 'use by' and 'best before' dates, which often cause confusion and lead to unnecessary food waste. By clarifying these labels, the EU aims to help consumers make better-informed decisions, thus reducing waste (EU Actions Against Food Waste, n.d.). The strategy also emphasizes the integration of food loss and waste prevention into broader EU policies and explores methods to prevent food losses at the production stage (EU Actions Against Food Waste, n.d.).

Building on the foundations laid by the Circular Economy Action Plan, the Green Deal also focuses on creating a common EU methodology to measure food waste consistently, facilitating food donation, and utilizing food no longer intended for human consumption in animal feed (Circular Economy Action Plan,2020). These initiatives are designed not only to minimize the environmental impact but also to address the economic and social implications of food waste. Reducing food waste contributes significantly to sustainability by lowering greenhouse gas emissions from food production and conserving resources used in food production, such as water and energy (Circular Economy Action Plan,2020). Overall, these comprehensive measures under the European Green Deal aim to enhance

the resilience and sustainability of the EU's food systems, addressing food waste at multiple stages of the supply chain and involving various stakeholders to foster a more circular and efficient food economy (Circular Economy Action Plan,2020).

The United Nations' (UN) Sustainable Development Goals (SDG) highlight the urgent need to address food waste through Goal 12: Ensure sustainable consumption and production patterns. Specifically, Target 12.3 aims to halve global food waste per capita at the retail and consumer levels and reduce food losses along production and supply chains by 2030 (Halving Food Waste and Raising Climate Ambition: SDG 12.3 and the Paris Agreement, n.d.). The UN Environment Program emphasizes that food loss and waste are significant global issues, with around one-third of all food produced being lost or wasted, which leads to enormous socio-economic and environmental costs. This inefficiency not only squanders resources but also exacerbates environmental degradation due to the need for additional agricultural land, impacting biodiversity and contributing to climate change (Halving Food Waste and Raising Climate Ambition: SDG 12.3 and the Paris Agreement, n.d.).

Reducing food waste not only addresses food security and economic efficiency but also plays a crucial role in mitigating climate change and preserving natural resources. As part of the global effort to implement the SDGs, coordinated actions involving governments, businesses, and consumers are essential to promote sustainable consumption and production patterns (Halving Food Waste and Raising Climate Ambition: SDG 12.3 and the Paris Agreement, n.d.).

The Scientific and Technological Research Council of Turkey has several initiatives and projects aimed at reducing food waste, which align with global efforts to address this critical issue. One notable example is their support for research and development projects focused on sustainable food systems and innovative technologies to minimize food loss and waste. For instance, TÜBİTAK funds projects that aim to improve food storage techniques, enhance supply chain logistics, and promote the use of data analytics to predict and manage food demand more accurately. Such initiatives are part of a broader strategy to ensure food security and sustainability in Turkey, in line with the UN’s SDGs, particularly with the SDG 12, which focuses on responsible consumption and production.

For instance, TÜBİTAK collaborates with international partners through joint calls such as the one organized by FOSC (Food Systems and Climate) and SUSFOOD2 (Sustainable Food Production and Consumption), supported within the Horizon 2020 Program. The main theme of this call is to find solutions to climate change, system shocks, and limited natural resources to make future food systems more sustainable. The projects supported under this call are expected to use resources efficiently, be resilient to climate effects, and offer innovative solutions. These projects are multidisciplinary or interdisciplinary and integrate various stakeholders, focusing on increasing resource efficiency and reducing waste while enhancing the adaptability and resilience of food systems to various shocks (*ERA-NET FOSC and SUSFOOD2 Common Call Opened | TÜBİTAK | Türkiye Bilimsel Ve Teknolojik Araştırma Kurumu*, 2024).

According to The UN SDGs on average, each individual wastes 121 kilograms of food per year (Home — SDG Indicators, n.d.). For this reason, demand prediction holds great significance. Demand prediction, a cornerstone of modern business strategy, involves predicting future consumer demand for products and services using historical data, statistical analysis, and machine learning techniques. It is an important issue for restaurants, markets or in general shops.

This thesis concentrates on the intricate domain of demand forecasting, exploring various methodologies, algorithms, and strategies aimed at enhancing both user and company experiences. The emergence of demand prediction has significantly transformed the way vendors develop their strategies, altering the process of estimating consumer demand for current goods and services, and aiding in the identification of what new products will capture attention. By accurately forecasting demand, businesses can optimize inventory levels, reduce expenses, enhance customer satisfaction, and improve overall operational efficiency. Furthermore, effective demand prediction enables businesses to be flexible and responsive in a dynamic market environment where customer preferences and market conditions are constantly evolving. This adaptability allows them to meet market demands while minimizing waste and maximizing revenue.

One of the objectives of this thesis is to contribute to the understanding and advancement of demand forecasting by conducting an extensive review and analysis of state-of-the-art

prediction algorithms and methodologies proposed specifically for the food sector. For this purpose, we first concentrated on an e-commerce market dataset, provided by a company in Turkey, which is a company producing creative technology solutions for various brands. This grocery shopping dataset consists of food names with unique food IDs, food categories, prices, weights and order quantities. Secondly, we worked on another dataset from Kaggle called Food Demand Forecasting, as it can be considered as a benchmark dataset in this domain. It has meal IDs, center and category information, price, location and number of orders as features. We encountered a paper using this dataset. Similar to our work they used Random Forest Regressor, Gradient Boosting Regressor, Light Gradient Boosting Machine Regressor, Extreme Gradient Boosting Regressor, Cat Boost Regressor, Long Short-Term Memory (LSTM), and Bidirectional LSTM (Panda & Mohanty, 2023).

Among the datasets we used both grocery shopping dataset and the food delivery business dataset exhibit similarities in seasonal variations, day-of-the-week patterns, time-of-day peaks, external influencing factors, and the impact of promotions. Seasonal trends affect both sectors, leading to higher demand during holidays and festive periods. Weekends typically see increased activity compared to weekdays, and both grocery stores and food delivery services experience peak times in the evenings, although grocery stores also see late afternoon surges. External factors such as weather, economic conditions, and local events play significant roles in shaping demand. Promotions and special offers are effective in driving demand for both groceries and food delivery services.

However, there are notable differences between the two. Grocery stores must manage a wide variety of products with different shelf lives, ranging from perishable to long-lasting items, necessitating careful inventory management to avoid spoilage and stockouts. In contrast, food delivery businesses focus on prepared meals with shorter shelf lives, requiring precise inventory control to ensure freshness and minimize waste. While grocery store demand is influenced by product variety and availability, food delivery demand hinges more on menu offerings, food quality, and the convenience of delivery. Customer decision-making also differs; grocery shoppers often follow predetermined lists but can be swayed by in-store promotions, whereas food delivery customers make convenience-driven choices based on menu options, delivery speed, and service quality. Additionally,

grocery shopping typically occurs once or twice a week with larger quantities purchased, whereas ordering food delivery is less frequent and varies by customer preferences and lifestyle. The data sources for predictive analytics also differ, with grocery stores relying on point-of-sale data and loyalty programs, while food delivery services utilize order history, delivery patterns, and customer preferences. In summary, while both sectors share some common demand patterns, their distinct operational and customer behavior characteristics may need tailored approaches to demand prediction.

The analysis and results will be explained thoroughly in the literature review section. On both dataset we trained seven machine learning models along with a Nonlinear Autoregressive with Exogenous Inputs (NARX) time-series model and we compared the results for these two datasets. The central focus lies in investigating demand forecasting approaches while striving to improve the model performance.

1.1. Contributions

The contributions of this thesis are outlined as follows:

- Exploration of the usage of various prediction models including five machine learning models and one neural network model,
- Performance comparison and analysis of different prediction models on two datasets with different characteristics,
- Giving insights for the model selection and its numerical application.

The rest of the thesis is organized as follows: in Section 2, we present a comprehensive literature survey of the research focusing on investigating the state-of-the art prediction models, seeking to identify their strengths and weaknesses in various domains. In Section 3, we describe and analyze the dataset and detailed definitions of the employed terms throughout the article. Next, the base methodology and our proposed framework are described in Section 4. An in-depth investigation was conducted on distinct demand forecasting algorithms. Section 5 presented the experiments and obtained results, and finally, in Section 6 we concluded the paper.

2. LITERATURE REVIEW

Predicting food demand utilizing learning-based models is important for sustainable development and business process optimization. We therefore investigated similar papers with our objective. One study in Poland aimed to create food demand prediction models using a nonlinear autoregressive exogenous neural network (NARXNN) (Lutoslawski et al., 2021). With the goal of decreasing food waste and loss, NARXNN, which is a special type of recurrent neural network providing better predictions than the traditional RNN was used. NARXNN has been a popular method for many fields like prediction of traffic condition, energy and water consumption, and soil moisture on hourly basis. In their study (Lutoslawski et al., 2021), the researchers worked on a dataset provided by a company that works on inventory management and demand prediction systems for producers and distributors. A time series containing demand values presented to the neural network during training constituted the training dataset. The testing dataset only included the time series for multi-step prediction testing and was not previously provided to the neural network. The model performance was analyzed based on the lowest values of the MAE, MAPE, and RMSE, followed by the greatest value of the determination coefficient (R-squared). They experimented with various delay line lengths, the number of hidden layers, and the number of neurons within the hidden layers to determine the optimal network architecture. Based on the study's findings, it is possible to conclude that very accurate food demand projections can be made using the artificial intelligence-based NARX models (Lutoslawski et al., 2021).

To anticipate the number of orders, another work used the Genpact-released "Food Demand Forecasting" dataset (Analytics Vidhya, 2020). This is also the dataset used in this paper. The researchers examined the impact of several factors on demand, extracted the typical traits that may have an influence, and suggested comparing seven regressors: Random Forest Regressor, Cat Boost Regressor, Extreme Gradient Boosting Regressor,

Gradient Boosting Regressor, Light Gradient Boosting Machine Regressor, Long Short-Term Memory, and Bidirectional LSTM in this study.

Another paper that used the "Food Demand Forecasting" dataset aimed to utilize historical data to forecast weekly grocery shipments for a specific business [34]. By analyzing historical demand patterns, industry trends, and other relevant sectoral factors, the business seeks to develop a robust forecasting model capable of accurately predicting future demand for food delivery. In this paper, data analysis was a crucial part and they visualized many aspects of the dataset. They used both one-hot encoding and label encoding for the categorical features. As a model, they also used RFR and they obtained MAE of 5.3 and RMSE of 7.2 proving that an accurate prediction is possible with RFR. In addition, they developed a web application for users with Flask and deployed their model [34].

A study in Turkey focused on this issue with the data from their refractory and they worked on developing machine learning-based prediction models designed to predict heavy demand for food during specific timeframes, such as lunch, without pre-booking (Aci & Yergök, 2023). They devised 18 prediction models employing five main techniques. These include three Artificial Neural Network models (Feed Forward, Function Fitting, and Cascade Forward), four Gaussian Process Regression models (Rational Quadratic, Squared Exponential, Matern 5/2, and Exponential), six Support Vector Regression models (Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian), three Regression Tree models (Fine, Medium, and Coarse), and two Ensemble Decision Tree (EDT) models (Boosted and Bagged), along with one Linear Regression model. This study stands out for its method diversity, prediction performance, and application scope, providing a distinctive contribution. The EDT Boosted model emerged as the top performer, achieving excellent prediction metrics, including a MSE of 0.51, MAE of 0.50, and an R-squared value of 0.96 (Aci & Yergök, 2023).

Retailers should pay special attention to demand prediction regarding fresh food and perishable goods supply chains. Since these types of foods spoil quickly, it is important to produce the right amount and acquire the necessary storage and delivery conditions. Underestimating or overestimating demand has a detrimental effect on the retailer’s

earnings. With this idea in mind, a decision support system was developed (Huber et al., 2017). They utilize multivariate autoregressive integrated moving average models for daily demand prediction on a dataset provided by the bakery chain for 18 months. They worked on a cluster-based hierarchical demand prediction system. The clusters were effective in lowering computing expenses without compromising the accuracy of the forecasts. Plus, they concluded that based on the cluster analysis, it is fair to estimate demand at an aggregate level because substitutable items have comparable intra-day sales patterns (Huber et al., 2017).

Since ARIMA is an important and popular model for demand prediction, another paper on this subject was analyzed (Nassar et al., 2020). A study in Canada focused on market price prediction by comparing deep learning and standard machine learning models. The models used in that paper were ARIMA, Support Vector Regression, Gradient Boosting, Artificial Neural Networks, LSTM, CNN, CNN-LSTM, Attention-CNN-LSTM. The models were applied to two datasets. The first dataset was taken from a website providing daily crop prices for Taiwanese marketplaces. The second dataset was taken from a confidential DC in Cambridge, Ontario, containing daily strawberry sales in the last seven years. In their paper, the researchers concluded that traditional machine learning models perform better than statistical models like ARIMA. Furthermore, the CNN-LSTM outperformed the basic ones like the LSTM. The model could accurately forecast prices up to three weeks in advance (Nassar et al., 2020).

In another work, a decision support system was developed specifically for food that goes bad quickly (Dellino et al., 2017). The researchers worked on a dataset that came from a real fresh food supply chain. It consists of 3-year sales, from 2011 to 2013, for a set of 19 small- and medium-sized supermarkets operating in Apulia region, Italy. The decision support system used ARIMA, ARIMAX, and transfer function prediction models. The decision support system they developed chooses the model that would work best based on the performance criteria given by the user. The model made sure prediction errors and computational effort are within reason (Dellino et al., 2017).

Demand prediction methods were analyzed within the context of the physical internet in (Punia et al., 2020). Researchers used hybrid approaches in response to the growing

diversity and shifting data flows. A novel strategy based on learning approaches was put forth considering the limitations of traditional prediction methods. RNNs with LSTM networks are especially well-suited for sequential data prediction applications like demand prediction and time series forecasting. The model LSTM-RNN was implemented with Keras and Sci-kit libraries in Python. Moreover, to compare the LSTM model, three other models were studied: Multiple Linear Regression, Support Vector Regression, and ARIMAX. They concluded that compared to the other three models, the LSTM model works better in terms of accuracy. In terms of the degree of association, LSTM was more useful to find the patterns of future demand based on the coefficient of determination (Punia et al., 2020).

In a paper on agricultural products, an LSTM is proposed for demand prediction in a physical internet supply chain network. A hybrid genetic algorithm and scatter search are proposed to automate tuning of the LSTM hyperparameters (Kantasa-Ard et al., 2020). Accuracy and coefficient of determination were the key performance indicators used to compare the performance of the proposed method with other supervised learnings: ARIMAX, SVR, and Multiple Linear Regression. The results proved the better prediction efficiency of the LSTM method with continuous fluctuating demand, whereas the others offer greater performance with less varied demand (Kantasa-Ard et al., 2020).

Above, we have provided the studies most similar to our thesis. In the literature, we have observed that various machine learning and time series models are frequently utilized in this field. For instance, the ARIMA model, SVR model, and Regression models, LSTM are among the most commonly employed methods. In our study, we utilized several regression and time-series models as the literature suggested. Similar to the study done in Poland (Kantasa-Ard et al., 2020), we give significant importance to NARX model, believing that it can be quite useful for food demand prediction, especially for fresh-food markets. Additionally, we also compared the results of our main model NARX with similar benchmark models to evaluate the performance of our proposed method. Our study differs in several key aspects from existing literature and methodologies: While many studies have focused on a certain dataset with various models, our research considers two

large datasets and compares the performances of various models on both datasets, highlighting the importance of data for model performance. The first dataset used in this study was provided by a company in Turkey and they collect their dataset from an online grocery store. It is an online grocery webpage based in Turkey. The dataset is local and unique, and its features are all in Turkish. This gave us an opportunity to analyze the customer behavior in Turkey and see the model performance in a completely different type of dataset. The other dataset is a public one and taken from Kaggle. This distinction not only sets our study apart but also contributes to a deeper understanding of customers

3. DATA ANALYSIS

This section explains two main datasets by outlining their features and the preprocessing procedures.

3.1 Online Market Dataset

At the outset of our study, we began examining this dataset to work on both demand prediction and recommendation systems. This dataset comprises seven years of sales data, spanning from January 2016 to September 2023, comprising a total of 2.699.563 order records with 11 columns in its raw form. The dataset encompasses a wide array of features concerning products and sales. The features and their explanations are as follow:

1. *OriginalId (Order Number):* It is an integer used to sort customer orders, and it remains the same for all products within the same order.
2. *Date (Date):* The date when the order was placed in the format Year-Month-Day Hour :Minute :Second.Millisecond
3. *FoodId (Food Identifier):* Unique integer values assigned according to the types of foods.
4. *ProductName (Product Name):* The Turkish name that uniquely identifies the product.
5. *OrderAmount (Order Quantity):* The quantity of the product ordered, either in units or kilograms.
6. *SelectedAmount (Selected Quantity):* The quantity of the product available in stock for the ordered amount, in units or kilograms according to the product.

7. *ActualWeight (Actual Weight):* The actual weight of the obtained product in units or kilograms according to the product.
8. *UnitPriceWithTax (Price with Tax):* The unit price of the product including taxes, in Turkish Lira, according to the product.
9. *UnitPrice (Price):* The unit price of the product excluding taxes, in Turkish Lira, according to the product.
10. *PickingType (Picking Type):* Products are categorized into six groups.
11. *EstimatedWeight (Estimated Weight):* The estimated weight of the product in grams.

	OriginalId	Date	FoodId	ProductName	ActualWeight	UnitPriceWithTax	PickingType	EstimatedWeight	Year	Month
0	30970	2016-01-01 11:43:26.000	99	Ispanak, Beypazarı (kg)	1.0	NaN	3	2500.0	2016	1
1	30970	2016-01-01 11:43:26.000	101	Karnabahar, Karacabey (adet)	1.0	NaN	2	NaN	2016	1
2	30970	2016-01-01 11:43:26.000	102	Kereviz (Yapraklı) , Geyve / Adapazarı , kg	1.0	NaN	3	1000.0	2016	1
3	30970	2016-01-01 11:43:26.000	103	Kıvırcık, Göçbeyliği (adet)	1.0	NaN	2	NaN	2016	1
4	30970	2016-01-01 11:43:26.000	104	Lahana - Beyaz, Samsun (adet)	1.0	NaN	2	NaN	2016	1

FIGURE 3.1.1 – Sample Data from the Online Market Dataset

However, we have specifically focused on a certain feature for our analysis, namely: Date, Food ID, Product Name, Picking Type, and OriginalID. Remaining variables were eliminated due to various reasons. Unit Price With Tax and Estimated Weight fields mostly have empty values which makes the variables useless for the model. On the other hand, Actual Weight field is not empty but most of its values are '1.0'. Therefore, it does not provide helpful insights to differentiate products. Date field was divided into Year and Month columns to put in the model. Product Name was used for the data description section to analyze and understand the data. Afterwards it was eliminated before the model, since we already have unique Food IDs to differentiate products. OriginalID was also used for descriptive data analysis but was not used in the model because it provided limited information. The Original ID field represents the customer orders. For all products in the same order, the Original ID is the same. Even though it gives us information about the order, it does not provide information about the customer behaviors. We cannot analyze the trends and patterns of a customer. Therefore, the recommendation system is left as a future work. These chosen attributes serve as valuable elements for both filtering data and creating predictions. Figure 3.1.1 illustrates a sample extracted from this dataset.

We first made a descriptive analysis on the data. The sale count was analyzed in terms of Year and Month to understand if any significant data would help improve the model. The sale count based on the Year and Month is in Figure 3.1.2. The *y*-axis shows the number of sales and *x*-axis shows the months. In the first four years the sales count was small. As we approached the present day, sales quantities increased, and we had a larger dataset. The graph in Figure 3.1.2 shows a trend in specific periods. Except for 2022, the sale was at its highest in the months April and May. In the years 2022, 2021, 2020 and 2016, the sale counts dropped significantly. This could be due to the holiday season. Moreover, we see a drastic increase in the February-March of 2020. This could be explained with the pandemic and the increase in online shopping due to long quarantines. As the virus began to spread globally, many countries implemented lockdowns, social distancing measures, and other restrictions to curb its transmission. These measures included closing non-essential businesses and encouraging or mandating people to stay at home. Consequently, consumers turned to online shopping to meet their needs, particularly for essential goods like groceries.

The panic buying and stockpiling of food and household items in anticipation of prolonged isolation further amplified this surge in demand. Traditional brick-and-mortar grocery stores faced challenges such as maintaining social distancing, managing increased foot traffic, and ensuring stock availability. Online grocery platforms, with their promise of home delivery and contactless transactions, became a safer and more convenient alternative for many. This sudden shift in consumer behavior resulted in a significant increase in online grocery orders, reflected in the sales data for February and March 2020.

Even after the initial surge in 2020, online shopping websites for groceries remained popular due to several factors that sustained and solidified this trend. First, the convenience and safety of having groceries delivered directly to one's doorstep continued to appeal to consumers, even as lockdowns eased, and brick-and-mortar stores reopened. Many people had adapted to the ease of online shopping and integrated it into their regular routines. Additionally, the improvements in online shopping platforms, such as better user interfaces, enhanced delivery logistics, and expanded product offerings, made the experience more attractive.

Retailers also invested heavily in digital infrastructure and services, including same-day delivery, subscription models, and personalized shopping experiences, which further entrenched online grocery shopping as a preferred option. Moreover, the ongoing health concerns and the

potential for future pandemics or similar disruptions kept the demand for contactless shopping high. The habits formed during the peak of the pandemic have had a lasting impact, leading to a continued preference for the convenience, safety, and efficiency provided by online grocery shopping websites.

This continued preference for online grocery shopping can be seen in the Figure 3.1.2 as well. From 2020 onwards, the data shows that the higher transaction counts continued into the following years, although with some fluctuations. The years 2021, 2022, and 2023 show sustained higher levels of online grocery shopping compared to pre-pandemic years (2016-2019). This proves that the pandemic accelerated the adoption of online grocery shopping, and the habit persisted even as the immediate impacts of the pandemic waned. The data highlights a significant behavioral shift in consumer purchasing patterns, with online grocery shopping becoming a more integral part of daily life post-2020. This increase enables us to work with a large dataset, thereby creating more accurate models.

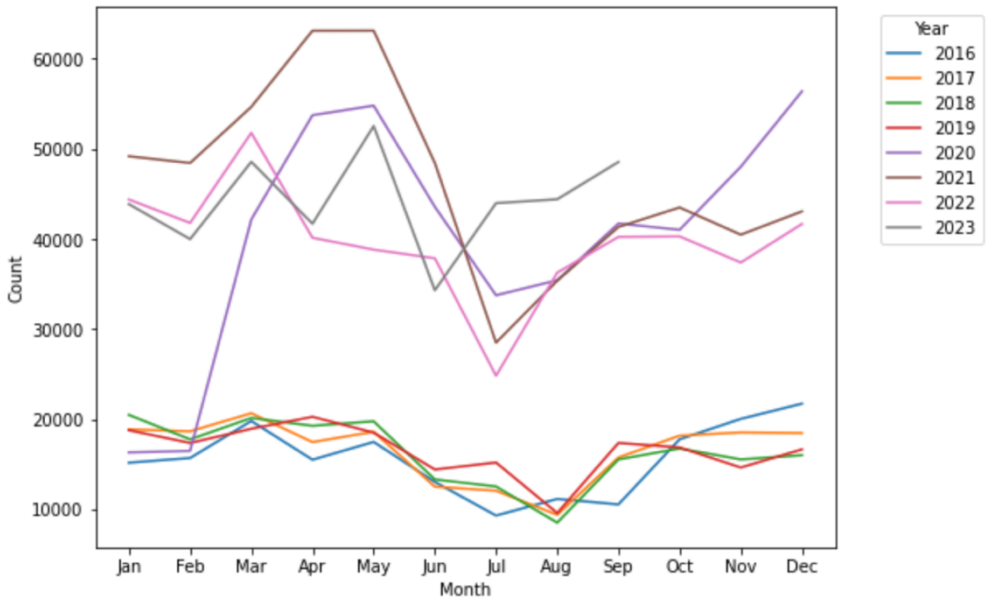

FIGURE 3.1.2 – Sale Count Based on Month and Year

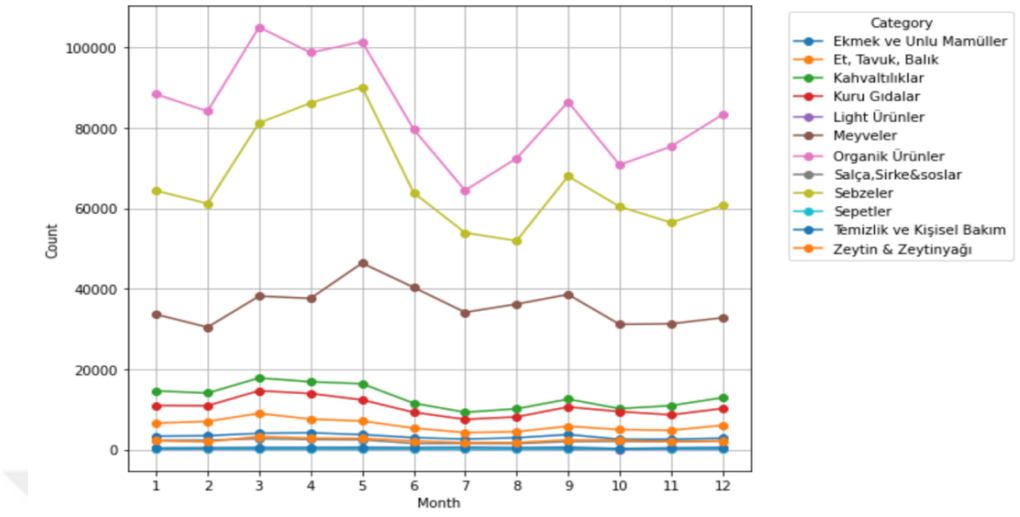

FIGURE 3.1.3 – Sale Count of Categories Based on Months

Figure 3.1.3 illustrates the sale count of categories based on month. As we stated above the sales was high in April and May, this is the same for categories as well. The most popular products are the organic products since for all the dates it has the highest number of sales. This is a large category, which constitutes the large portion of the data. Cleaning and personal care products have the smallest number of sales and follow a stable trend, as opposed to the organic products and vegetables categories, in which the values fluctuate. One of the primary conclusions we can draw is that the demand for vegetables and organic products during the summer season is significantly lower. This observation is crucial for inventory management, especially given that both categories consist of perishable goods. The perishability of these items implies a higher risk of waste if supply does not match the real-time demand. The data, which is limited to Istanbul, shows a noticeable drop in sales for organic products, fruits, and vegetables during the summer months. This trend can be attributed to the seasonal migration patterns of the city's residents, who often leave for vacation homes or holiday destinations. As people spend time away from Istanbul, they are less likely to order from online markets, leading to a decline in local demand.

This seasonal fluctuation necessitates strategic adjustments by market suppliers. By anticipating the reduced demand during the summer, markets can decrease their stock of perishable

goods accordingly. This proactive approach can prevent overstocking and minimize the financial losses associated with spoilage. For example, reducing the inventory of organic products, fruits, and vegetables to align with the expected lower demand can significantly decrease waste. Furthermore, understanding these seasonal trends enables markets to optimize their supply chains. By forecasting demand more accurately, they can plan for procurement, storage, and distribution more effectively. This ensures that during the high-demand periods, there is adequate stock to meet consumer needs, while during the low-demand summer season, the supply is scaled back to prevent excess.

In conclusion, aligning inventory levels with seasonal demand patterns is essential for managing perishable goods. The observed decrease in sales during the summer months in Istanbul highlights the importance of market adaptation. By reducing stock in response to lower demand, markets can decrease waste, improve efficiency, and ultimately enhance profitability. This strategy not only addresses the challenges posed by perishable goods but also supports sustainable business practices by minimizing food waste.

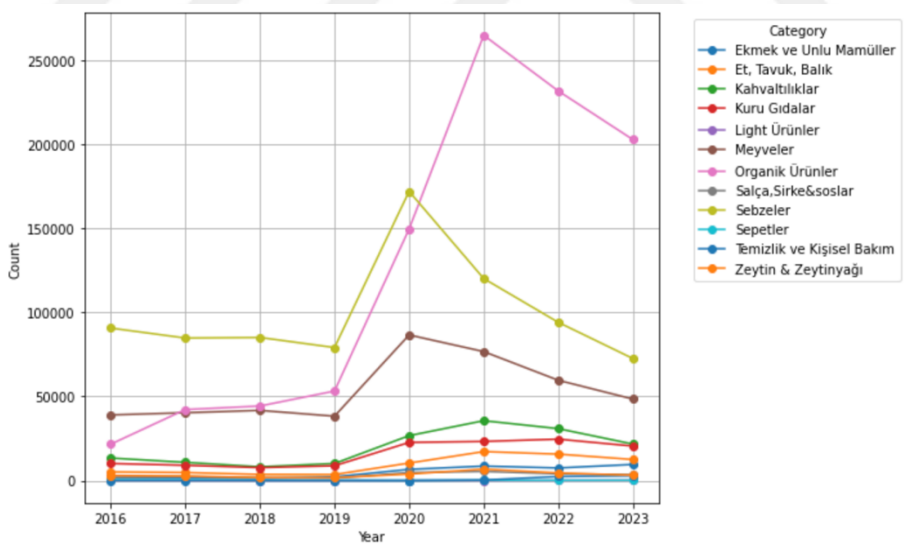

FIGURE 3.1.4 – Sale Count of Categories Based on Years

Figure 3.1.4 illustrates the yearly count of various product categories from 2016 to 2023. It encompasses twelve distinct categories, each denoted by a unique color and line style. The categories are Ekmek ve Unlu Mamüller (Bread and Bakery Products), Et, Tavuk, Balık (Meat, Chicken, Fish), Kahvaltılıklar (Breakfast Items), Kuru Gıdalar (Dry Foods), Light Ürünler (Light Products), Meyveler (Fruits), Organik Ürünler (Organic Products), Salça, Sirke & Soslar (Paste, Vinegar & Sauces), Sebzeler (Vegetables), Sepetler (Baskets), Temizlik ve Kişisel

Bakım (Cleaning and Personal Care), and Zeytin & Zeytinyağı (Olives & Olive Oil). Among these categories, 'Light Ürünler' exhibits the most significant variability, showing a dramatic increase starting in 2019, peaking in 2021, and then gradually declining by 2023. 'Organik Ürünler' also shows a substantial rise beginning in 2019, reaching a peak in 2020, followed by a slight decrease and stabilization. Categories such as 'Kuru Gıdalar' and 'Temizlik ve Kişisel Bakım' display a more consistent upward trend with moderate increases throughout the observed period. Other categories like 'Meyveler,' 'Sebzeler,' and 'Kahvaltılıklar' remain relatively stable with minor fluctuations. The data suggest a noteworthy increase in certain product categories, particularly 'Light Ürünler' and 'Organik Ürünler,' around 2019-2021, reflecting possibly changing consumer preferences or market trends during those years. The year 2023 shows a continuation of the downward trend for 'Light Ürünler,' while other categories maintain their trajectories.

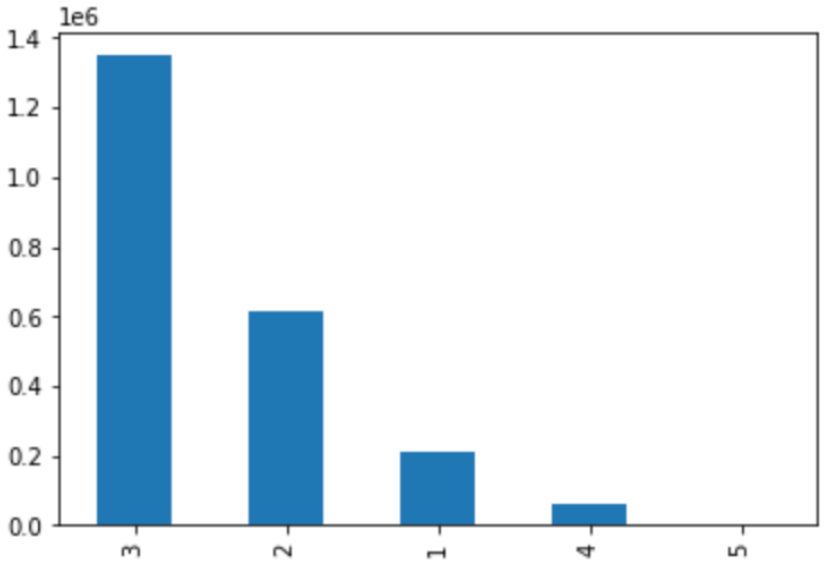

FIGURE 3.1.5 – Sale Count of Picking Types

The bar graph illustrates the distribution of various picking types based on their count. The feature name “picking type” is given by the company that created the dataset. This field can be considered as a category field with six different values. The *x*-axis represents different categories of products, each corresponding to a specific picking type, while the y-axis represents the count of items in each category. The categories are:

1. "Paketli ürün" (Pre-packaged product)
2. "Tartılmadan adetle satılan ürün" (Products sold by count without weighing)
3. "Tartılarak satılan çok adetli ürünler" (Products sold by weight in bulk)

4. "Tartılarak satılan tek adet ürünler" (Products sold by weight individually)
5. "Kahvaltılık ürünler" (Breakfast products)
6. "Et ürünleri" (Meat products)

From the graph, it is evident that "Tartılarak satılan çok adetli ürünler" (Products sold by weight in bulk) has the highest count, reaching nearly 1.4 million. This is followed by "Tartılmadan adetle satılan ürün" (Products sold by count without weighing) with around 700,000 and "Paketli ürün" (Pre-packaged product) with about 500,000. The counts for "Tartılarak satılan tek adet ürünler" (Products sold by weight individually) and "Kahvaltılık ürünler" (Breakfast products), are significantly lower, with "Tartılarak satılan tek adet ürünler" being the highest among them but still below 100,000. “Et ürünleri” (Meat products) have zero records between the selected time intervals. This suggests that bulk products sold by weight are the most common, whereas individual and specialized product categories like breakfast and meat items are less frequently encountered.

Lastly, Figure 3.1.6 illustrates the sale count for each category. The above 3 categories have very few numbers of sales in our dataset, therefore the bar graph shows them very close to zero. Top 3 categories are Fruits, Vegetables and Organic Products. This is expected since this is considered as a very large group with many different products. For instance organic products has organic vegetables, organic fruits and even organic cleaning supplies. Even though these items could also belong to other categories, since they are organic, they are represented in the organic products category. For example, an item named organic apple would belong to the organic products even though it is actually a vegetable. Since the category field is important this may create confusions for the model.

Moreover, most of the dataset is perishable goods like fruits, vegetables, meat, chicken, fish and organic products. Perishable goods, like these, present unique challenges for demand prediction due to their limited shelf life. These items need to be sold within a short time frame to avoid spoilage and degradation in quality. This short shelf life makes demand forecasting more complex, as it requires a high level of accuracy to balance supply and demand effectively. Seasonal variations and trends significantly influence the demand for perishable goods. For instance, certain fruits may be in higher demand during summer months, or specific baked goods might see a surge in sales during holiday seasons. Additionally, weather conditions can impact both the supply and demand sides, adding another layer of complexity. Effective inventory

management is crucial; overstocking can lead to significant waste and increased costs, while understocking can result in missed sales opportunities and customer dissatisfaction.

In contrast, non-perishable goods, which include items like canned foods, household supplies, and electronics, have a much longer shelf life, allowing them to be stored for extended periods without the risk of spoilage. This extended shelf life simplifies inventory management compared to perishable goods, as businesses can afford to maintain larger stock levels without the immediate pressure of potential waste. However, demand prediction for non-perishable goods still requires careful consideration of various factors. Consumer preferences, economic conditions, and promotional activities can all influence demand. Additionally, non-perishable goods can exhibit cyclical demand patterns, such as higher sales of canned goods during winter months or increased demand for school supplies at the start of the academic year. While the stakes may not be as high in terms of waste, inaccurate demand forecasts can still lead to overstocking, tying up capital, or understocking, resulting in stockouts and lost sales.

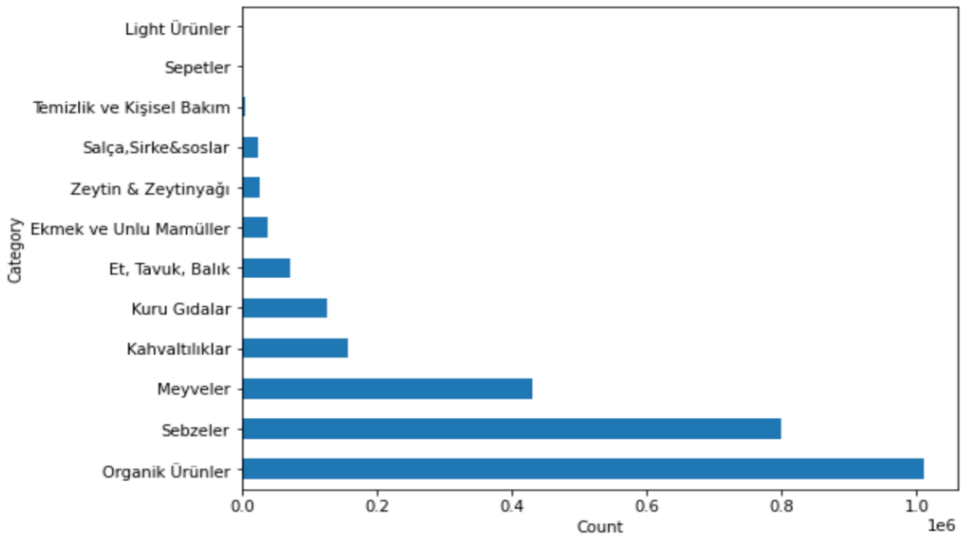

FIGURE 3.1.6 – Count of Each Category

Table 3.1.1: Statistics of Sales for Each Category

Category	Maximum Sales Date	Maximum Sales	Mean Sales
Ekmek ve Unlu Mamüller	8.02.2023	12	1,27
Et, Tavuk, Balık	28.03.2020	14	1,47

Kahvaltılıklar	25.05.2020	18	1,65
Kuru Gıdalar	3.02.2021	21	1,70
Light Ürünler	5.08.2023	5	1,17
Meyveler	28.07.2023	21	2,70
Organik Ürünler	28.03.2020	44	5,21
Salça, Sirke & Soslar	22.07.2020	8	1,25
Sebzeler	13.01.2016	47	4,85
Sepetler	28.02.2016	2	1,01
Temizlik ve Kişisel Bakım	23.06.2023	16	1,56
Zeytin & Zeytinyağı	14.01.2021	6	1,19

The table presents a detailed analysis of sales performance across various product categories, highlighting key metrics such as maximum sales, mean sales, and the dates of peak sales. The dataset spans different product types ranging from bread and bakery products (Ekmek ve Unlu Mamüller) to fruits (Meyveler), organic products (Organik Ürünler), and cleaning supplies (Temizlik ve Kişisel Bakım). This data offers valuable insights into consumer behavior and market trends.

Firstly, the maximum sales figures range widely across categories, from as low as 2 units to a high of 47 units. The category with the highest maximum sales is Sebzeler (Vegetables) with 47 units, recorded on January 13, 2016. In contrast, the lowest maximum sales are observed in Sepetler (Baskets) and Light Ürünler (Light Products) with only 2 and 5 units respectively. As explained before, vegetables category is a large category, so it is expected that it has a large number of sales in a day. The same thing is valid for Organic Products as well. For instance, The highest mean sales are found in Organic Products and Fruits with 5,21 and 2,70 units respectively, highlighting strong consumer interest in these categories. However, Baskets are not sold very often since they are only available in certain times. Such variations underscore the diverse demand levels and popularity of different product types. Each category's peak sales date provides insight into when consumers typically purchase these products in higher quantities, which could be influenced by seasonal factors or promotional activities.

Secondly, examining the dates of maximum sales provides context-specific insights. For example, Organic Products achieved its highest sales of 44 units on March 28, 2020. This peak

coincided with the onset of the COVID-19 pandemic, suggesting a possible increase in consumer preference for organic and healthier options during times of health crises.

Furthermore, the mean sales figures offer a glimpse into sustained consumer interest over time within each category. The mean sales per category also vary, indicating different levels of consumer demand and market penetration. Categories like Fruits and Kuru Gıdalar (Dry Foods) show relatively high mean sales figures, indicating consistent demand throughout the periods covered by the dataset. On the other hand, Fruits category is a very large category. Each fruit has its own season and knowing that there is a consistent demand for the general fruit category may not help us very much. We may need more information about this category.

Additionally, the table reveals some categories with notably low maximum sales, such as Zeytin & Zeytinyağı (Olives & Olive Oil) with a maximum of 6 units. Understanding such trends can be crucial for inventory management and marketing strategies tailored to boost sales in underperforming categories.

As stated before, this dataset was taken from a real market in Turkey. The company needs to predict the demand specifically for perishable goods because they were buying them from the wholesale market. Like demand prediction, a company like this may need supply prediction as well. Demand prediction is the process of forecasting future customer demand for products. It involves analyzing historical sales data to predict how much of each product customers will order in the future. This prediction helps businesses manage inventory, optimize supply chain operations, and ensure they meet customer demand without overstocking or understocking products. Supply prediction, on the other hand, involves forecasting the future availability of products. This prediction is crucial for managing inventory levels, planning restocks, and ensuring the supply chain can meet the predicted demand. The objective is to predict how much of each product will be available to fulfill future orders. In this dataset, attributes such as SelectedAmount, ActualWeight, EstimatedWeight, and PickingType provide valuable insights into the supply side. The SelectedAmount indicates the quantity available in stock, while ActualWeight and EstimatedWeight can be used to predict variations in product availability due to weight discrepancies. By analyzing these features, you can develop models that forecast future stock levels, considering factors like supplier lead times, stock replenishment rates, and inventory turnover. This helps in anticipating potential shortages or surpluses and adjusting procurement plans accordingly.

Both demand and supply prediction models rely on historical data to make forecasts but focus on different aspects of the business. Demand prediction is customer-centric, focusing on predicting customer orders and sales trends. It uses data related to customer orders, prices, and product characteristics to forecast future demand. In contrast, supply prediction is supply chain-centric, focusing on predicting product availability. It uses data related to stock levels, product weights, and supply chain processes to forecast future stock availability. Despite their different focuses, both types of predictions are interrelated. Accurate demand predictions inform supply predictions by indicating how much stock is needed to meet future demand. Conversely, supply predictions inform demand predictions by highlighting potential supply constraints that might affect future sales.

This dataset contains rich information for both demand and supply prediction. By leveraging attributes like order quantities, stock levels, product weights, and dates, you can build robust models to forecast future demand and supply. These predictions help optimize inventory management, reduce costs, and improve customer satisfaction by ensuring that products are available when and where they are needed. Therefore, as future work, this dataset could also be useful for supply prediction as well.

In conclusion, this detailed dataset not only provides quantitative metrics but also prompts further qualitative analysis, such as the impact of external factors like seasons, economic conditions, and societal trends on consumer purchasing behaviors across diverse product categories. Such insights are invaluable for businesses aiming to optimize their product offerings and capitalize on market opportunities effectively.

3.2 Food Demand Forecasting Dataset from Kaggle

This dataset was collected from a food delivery business with locations in several cities (Analytics Vidhya, 2020). For shipping meal orders to their clients, they have several fulfillment centers in these cities. The data comes in four different files: Train, test, meal information, and fulfillment center information. For this thesis, the train data was merged with meal information and fulfillment center information. The merging operation was done with the unique IDs. Train dataframe was merged with the fulfillment center information dataframe on center ID and it was merged with meal information dataframe on meal ID. This way we obtained one large dataset consisting of all the necessary information to run in a model. The dataset comprises 145

weeks of weekly orders for 51 distinct meals and 77 fulfillment centers. This data is distributed across three types of centers: Center A, Center B, and Center C, totaling approximately 456.548 entries and encompassing 14 features. Some additional features like checkout price, base price promotion method, operation area, city and region code were added. The features of the merged dataset are represented in Table 3.2.1 along with the feature definitions.

Table 3.2.1: Features of Food Demand Forecasting Merged Dataset

Features	**Description**
ID	unique order ID associated
week	week number
meal_id	unique ID associated with each meal
center_id	unique ID associated with each fulfillment center
checkout_price	final price of the meal
base_price	base price of the meal as written in the menu
emailer_for_promotion	whether an e-mail is sent for the promotion (0/1 - no/yes)
op_area	area of service for each center
city_code	pin code associated with each city
center_type	type of the center (type_A, type_B, type_C)
region_code	unique code associated with each region
category	type of meal
cuisine	Indian, Italian, etc

In the context of a food delivery service, the distinction between perishable and non-perishable goods becomes particularly important for demand prediction due to the unique nature of the business. Food delivery services deal with a variety of products, ranging from fresh meals and ingredients to packaged snacks and beverages.

Perishable goods in a food delivery service include fresh meals, salads, dairy products, fruits, vegetables, and any other items that require refrigeration or have a short shelf life. The demand for these items is highly variable and can be influenced by factors such as time of day, day of the week, weather conditions, and special events or holidays. For example, there may be a higher demand for fresh salads and cold beverages during hot weather, while demand for hot meals may increase during colder months.

One of the main challenges in predicting demand for perishable goods in food delivery is managing the short shelf life of these products. Accurate demand forecasting is critical to minimize waste and ensure that fresh items are available to meet customer orders. This requires sophisticated forecasting models that can account for seasonality, trends, and real-time data such as current weather conditions and local events. Additionally, logistical considerations such as delivery time and storage conditions play a significant role in maintaining the quality of perishable goods.

Non-perishable goods in a food delivery service include items like canned foods, bottled beverages, dry snacks, and other shelf-stable products. These items have a longer shelf life and can be stored without immediate risk of spoilage. Demand for non-perishable goods is generally more stable and less affected by immediate external factors compared to perishable goods. However, they can still be influenced by promotions, marketing campaigns, and changes in consumer preferences.

For non-perishable goods, the focus of demand prediction is often on optimizing inventory levels to ensure that popular items are always in stock while avoiding excessive inventory that ties up capital. Since these items can be stored for longer periods, businesses have more flexibility in managing their stock. Additionally, non-perishable foods can serve as a buffer to meet unexpected spikes in demand, providing a reliable supply for customers even when demand forecasting is not perfectly accurate.

In addition, the histogram of the features is illustrated in Figure 3.2.1. The distribution of weeks is fairly uniform, with a slight increase in frequency toward the later weeks. This suggests consistent sale over time, with a slight increase in data entries as time progresses. The center ID histogram shows two distinct peaks, indicating that certain centers (those around IDs 50-60 and 110-130) have significantly more entries than others. This could imply that these centers are

larger, more active, or more prominent in our dataset. The meal ID distribution is multimodal, with several peaks, suggesting that certain meal IDs are much more common. This indicates a few popular meals that are ordered more frequently compared to others. Category histogram shows a varied distribution with multiple peaks, indicating a diverse set of categories. Some categories are more prevalent than others, suggesting a mix of popular and less common categories. In summary, our dataset appears to be diverse, with certain prominent centers, meals, and categories. The data is collected consistently over time, with a few high-frequency centers and meals suggesting possible focal points or popular items. The distributions provide valuable insights into operational dynamics, customer preferences, and the overall structure of the dataset.

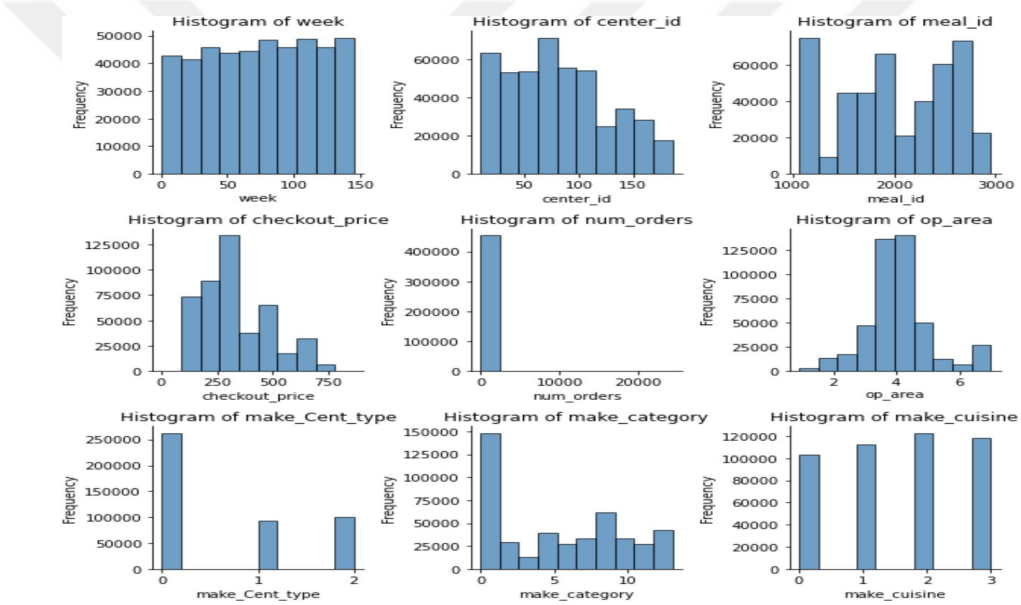

FIGURE 3.2.1 – The Histogram of Each Feature in the Dataset

Figure 3.2.2 illustrates the number of meal orders that each center type got in the dataset. The most meal orders were given to Type_A centers, while the fewest were given to Type_C centers. It is expected, since the number of Type_A centers is the highest. Figure 3.2.3 shows the distribution of meal orders based on the meal category. The beverage category has a surprisingly high number of sales, whereas the food category with the fewest orders is biryani. For having a general idea about the demands, the top five meals ordered are examined in Figure 3.2.4. The meal id with the highest number of orders is 2290 with around 9 million sales. Figure 3.2.5 illustrates the number of meals sold per week. This information is useful to understand the times that the demands are high or low. The weeks where the number of orders is high can be a specific date or holiday like New Years. By examining those weeks, the companies can be prepared in advance.

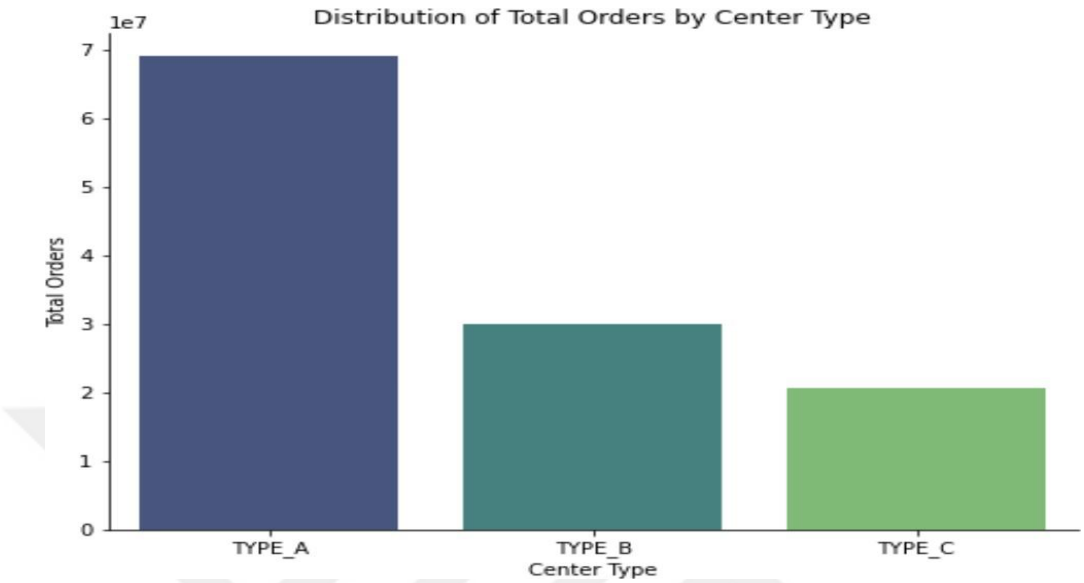

FIGURE 3.2.2 – The Number of Meals Sold per Center Type

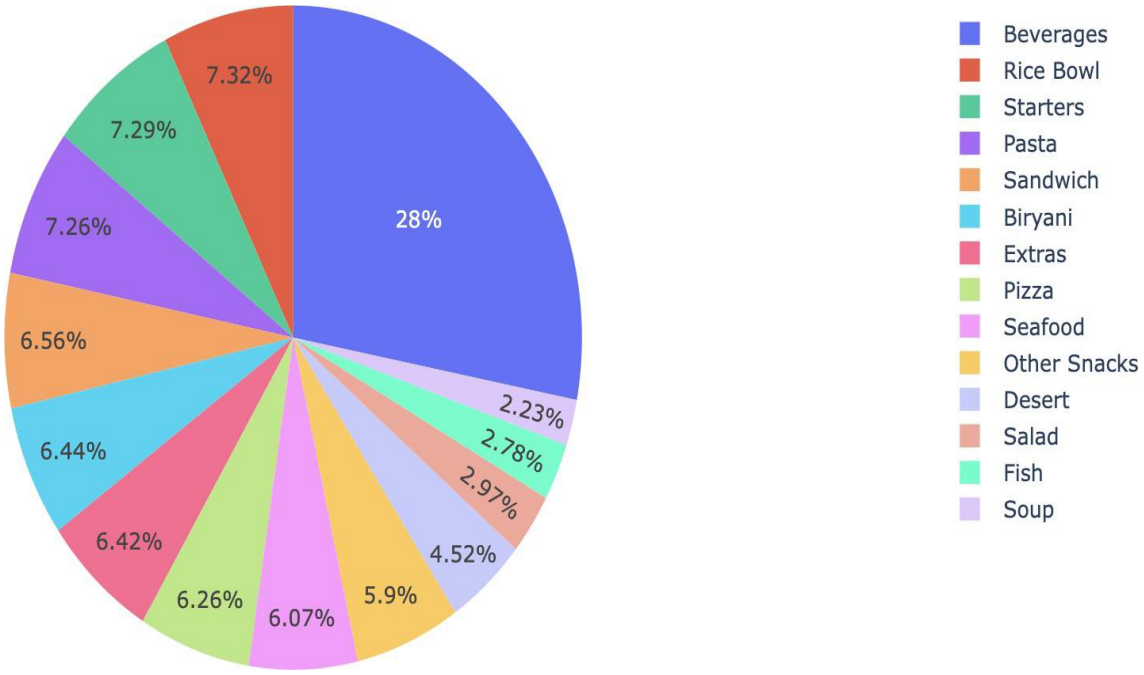

FIGURE 3.2.3 – The Number of Meals Sold per Food Category

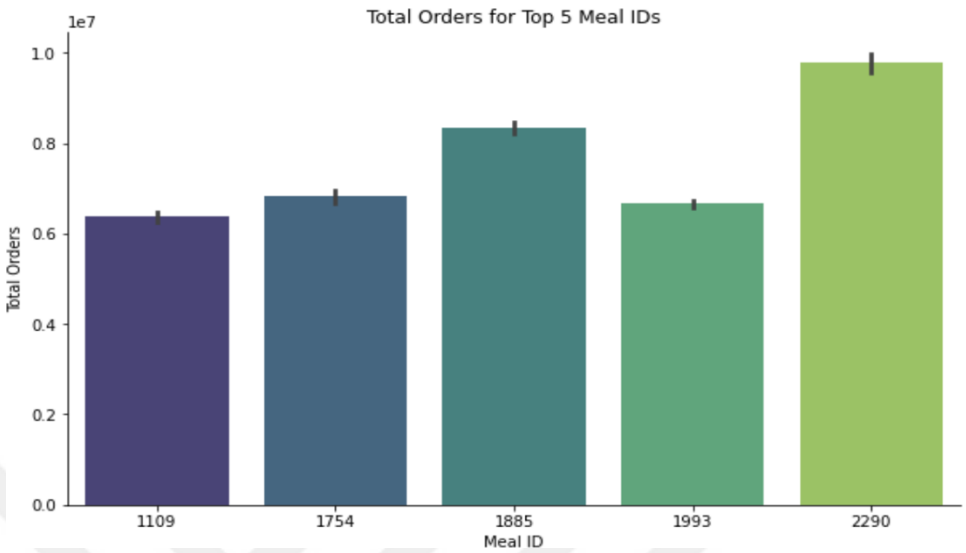

FIGURE 3.2.4 – Top 5 Meals Sold with Meal ID

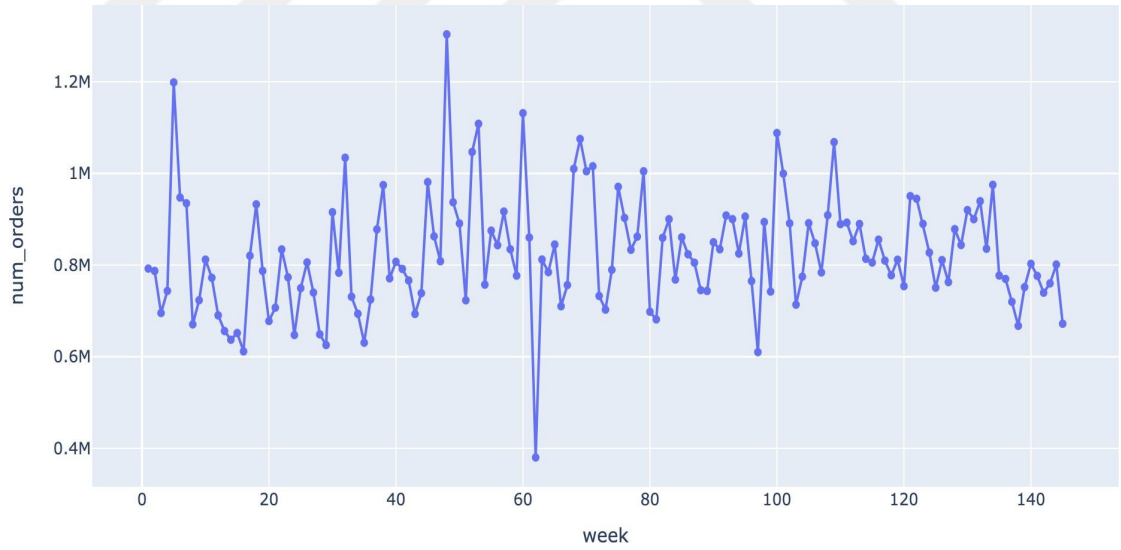

FIGURE 3.2.5 – The Number of Meals Sold per Week

In addition, Figure 3.2.6 shows the correlation between the features. The correlation matrix presented reveals several noteworthy relationships among the features in the dataset. Most prominently, there is a very high positive correlation between base price and checkout price (0,95), indicating that these two features are closely linked, which is expected considering that base price influence the final checkout price. Due to this high correlation,

one of the variables can be eliminated from the model. Additionally, a moderate positive correlation (0,39) exists between being featured on the homepage and being promoted via email, suggesting these promotional strategies often coincide. The number of orders shows moderate positive correlations with both homepage features (0,29) and e-mail promotions (0,28), implying that these marketing efforts effectively drive sales and one of these variables could be useful for model building. Furthermore, the make category has a moderate positive correlation (0,28) with checkout price, suggesting that specific categories tend to be priced higher. Similarly, there is a moderate positive correlation (0,25) between meal ID and checkout price, indicating that certain meals are more expensive. The relationship between make cuisine and make category is also noteworthy, with a moderate positive correlation (0,13), suggesting a link between the type of cuisine and its category. While most other correlations are weak, highlighting minimal linear relationships, these key insights help identify which features have strong associations, guiding further analysis and predictive modeling efforts. The matrix underscores areas where changes in one variable could predict changes in another, aiding strategic decision-making.

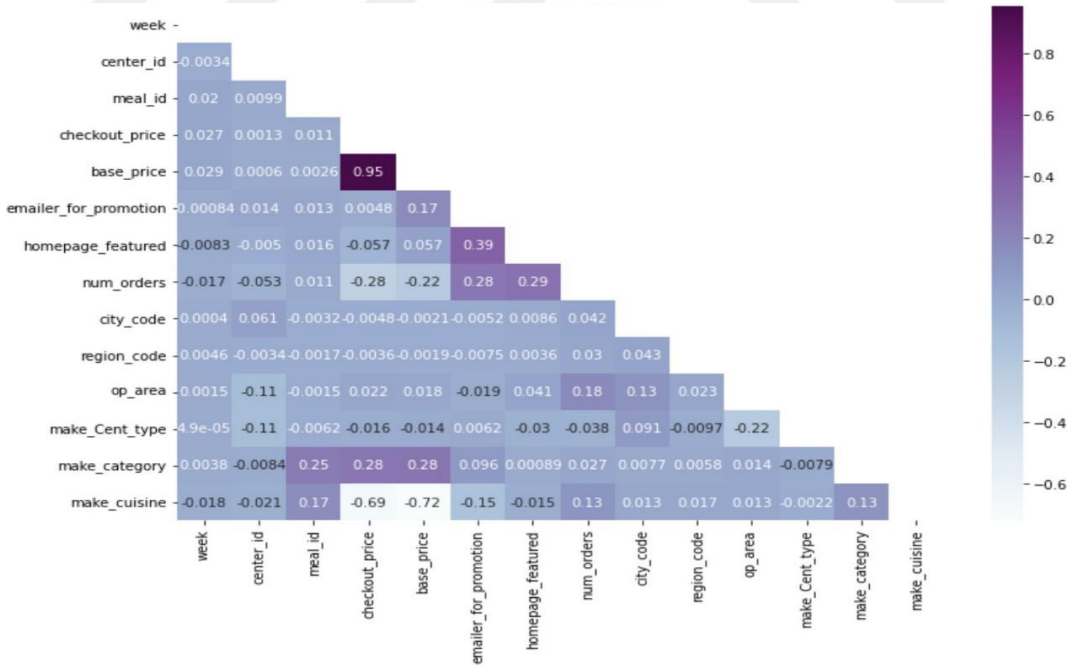

FIGURE 3.2.6 – The Correlation of Features

The scatter plot in figure 3.2.7 illustrates the relationship between checkout prices and the number of orders. The x-axis represents the checkout price, ranging from 0 to 800, while the y-axis indicates the number of orders, extending up to 25.000 units. The plot is densely populated at lower checkout prices, particularly between 50 and 200 units, showing a high concentration of orders in this price range. There is a significant cluster of points around this range, indicating that most orders are placed at these lower prices. There is a negative correlation between the checkout price and the number of orders. As the checkout price increases, the number of orders decreases, with a noticeable decline beyond the 200-unit mark. This makes sense because as the price increases, the demand decreases. There are a few outliers, especially one with a checkout price around 200 units and an exceptionally high number of orders nearing 25,000. The data points become increasingly sparse as the checkout price moves beyond 300 units, indicating fewer orders at higher prices. Overall, the scatter plot reveals a clear trend where lower checkout prices correspond to higher order volumes.

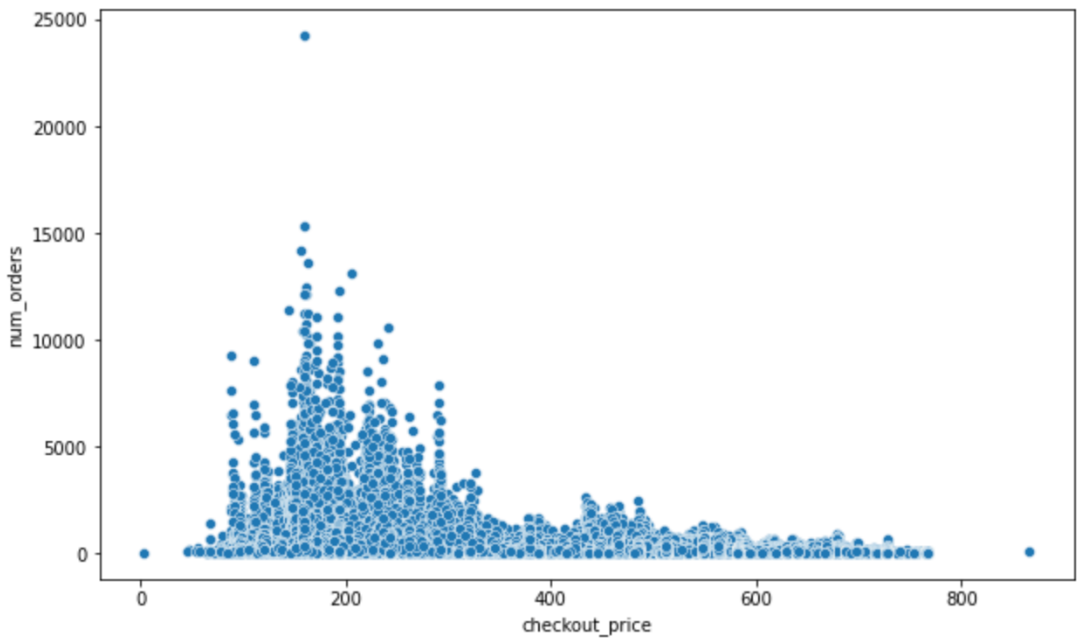

FIGURE 3.2.7 – Checkout Price vs Number of Orders

We conducted a thorough analysis of the provided data using various plots and graphs. This comprehensive examination not only enhanced the understanding of the data but also provided robust support for the conclusions drawn. The correlations plot in Figure

3.2.6 helped us determine which features to use in our dataset. For instance, we can conclude that base price and checkout price features are highly correlated, and one can be eliminated from the model. By looking at Figure 3.2.3, we can see that the dataset has many categories. However, one disadvantage we realized was that the beverages category makes up a significant portion of the data. On the other hand, some categories are considerably small which may cause problems to the model for the prediction task.

3.3 Dataset Comparisons

The two datasets, while both related to orders and products/meals, present a mix of similarities and differences that highlight their unique purposes and the scope of the data they capture. Both datasets include unique identifiers for the items within the orders, with the first dataset using *FoodId* for food products and the second using *meal_id* for meals, which is essential for detailed item-level analysis.

One of the key similarities between the datasets is their inclusion of pricing information. The first dataset offers detailed price attributes, including *UnitPriceWithTax* and *UnitPrice*, while the second dataset provides *checkout_price* and *base_price*. This pricing information is crucial for financial analysis and understanding revenue generation from each order. Furthermore, both datasets categorize the items, though in different ways: the first dataset uses *PickingType* to categorize products, while the second dataset categorizes meals by category and cuisine. This categorization allows for segmentation analysis, helping to identify trends and preferences in different product or meal categories.

Temporal data is another common feature, though it is represented differently in each dataset. The first dataset uses a detailed date field, capturing the exact date and time of the order, which is valuable for time series analysis and understanding order patterns over time. The second dataset uses a week field, indicating the week number, which is useful for aggregating orders over a broader time period and identifying weekly trends.

However, the datasets also exhibit significant differences. The first dataset provides a detailed structure of each order, including attributes such as *OrderAmount*, *SelectedAmount*, and *ActualWeight*. This granularity allows for in-depth analysis of each order's

composition, inventory management, and logistics. In contrast, the second dataset is more aggregated, focusing on the overall attributes of the orders and meals, and includes contextual information such as *center_id*, *center_type*, *op_area*, *city_code*, and *region_code*. This context is valuable for analyzing the broader operational environment, including the performance of different fulfillment centers and geographical demand patterns.

Another notable difference is the presence of promotional information in the second dataset, with the *emailer_for_promotion* feature indicating whether a promotional email was sent. This information is crucial for evaluating the impact of marketing efforts on demand. In the correlation table we have seen that there is a moderate correlation between the number of sales and the promotion methods. Moreover, as we will explain the results section, the NARX model had better results when we add the promotional features to the model. So, we can say that these features in the second dataset are really important and effective. Additionally, the second dataset includes detailed geographical information, providing insights into regional demand and service areas, which is absent in the first dataset.

The first dataset also includes detailed weight information (*EstimatedWeight* and *ActualWeight*), which is not present in the second dataset. This weight data is essential for logistics and inventory management, allowing for precise tracking of product quantities.

A particularly important aspect to note is the geographical and contextual differences between the two datasets. The first dataset is from an online grocery shopping site in Turkey, while the second dataset is from a food delivery business abroad. This distinction is crucial as it provides insights into demand patterns and operational dynamics in different regions and cultural contexts. Working with a real-life dataset from Turkey allows for the exploration of local market trends, consumer behaviors, and logistical challenges specific to the region. It is especially valuable for businesses operating in Turkey or looking to enter the Turkish market, providing a deep understanding of local demand.

On the other hand, the dataset from abroad offers a comparative perspective, highlighting differences and similarities in demand prediction across different geographical and cultural settings. This comparison is beneficial for developing generalized models that can

be adapted to various markets and for understanding how local nuances affect demand prediction.

Moreover, the type of food in these datasets differs significantly. The first dataset, being from an online grocery shopping site, includes a variety of grocery products such as raw ingredients, packaged foods, and household items, while the second dataset from a food delivery business likely includes prepared meals and ready-to-eat items. This variation in product type affects demand patterns and the factors influencing demand, such as preparation time, shelf life, and consumer purchasing behavior.

In a food delivery service, the complexity of demand prediction is heightened by the diverse range of items that make up a meal. Unlike retail environments where products are sold individually, food delivery involves meals composed of multiple components, each with its own demand characteristics and shelf life. This necessitates a more granular approach to demand prediction, where each component—such as salt, sauce, meat, and fries—must be analyzed separately and in conjunction with others.

When predicting demand for a food delivery service, it's essential to consider the intricacies of each meal component. For instance, a simple meal like a burger with fries and a drink involves multiple perishable and non-perishable items. Items like fresh meat, lettuce, tomatoes, and bread are perishable and require precise demand forecasting to ensure freshness and minimize waste. These items have a short shelf life and are susceptible to spoilage if not managed properly. Items such as salt, packaged sauces, and canned beverages have a longer shelf life and can be stored for extended periods. These items provide more flexibility in inventory management but still need accurate demand forecasting to prevent stockouts or overstocking.

To effectively manage a food delivery service, demand forecasting must integrate both perishable and non-perishable components. As opposed to the online market data for food delivery service dataset, each item within a meal should be forecasted individually. This means predicting the demand for fresh ingredients like meat and vegetables separately from non-perishable items like condiments and drinks. Historical sales data, seasonal trends, and promotional activities should all be factored into these forecasts.

Understanding how different components come together to form a meal is crucial. For example, if a particular burger is popular, the demand for its individual components—buns, patties, lettuce, and sauces—will increase accordingly. Analyzing sales data at the meal level helps in forecasting the demand for each component accurately. Incorporating real-time data such as current weather conditions, local events, and even time of day can significantly enhance demand prediction accuracy. For instance, demand for cold beverages may spike on hot days, while hot meals might be preferred during colder weather.

Effective demand prediction must be paired with efficient inventory management and supply chain coordination. This ensures that perishable items are procured and used within their shelf life, while non-perishable items are stocked in optimal quantities to meet demand without tying up excessive capital. For this thesis we don't have real-time information like current weather, local events or time of the day. In addition, we don't have the names of meals, we only have their ids and categories, so it is not possible for us the analyze the components of a meal individually. With our work we can only predict the demand for a specific meal but as a future work, with a more detailed dataset more detailed demand prediction could be achieved.

In summary, while both datasets aim to capture order and product/meal information, they do so from different perspectives and levels of detail. The first dataset's granularity is valuable for operational and logistical analysis, providing detailed insights into each order's composition and inventory needs. The second dataset's contextual information supports strategic decisions related to market analysis, regional demand patterns, and the effectiveness of promotional activities. Integrating insights from both datasets offers a comprehensive view of demand patterns, optimizing both operational and strategic decision-making processes. The inclusion of a local Turkish dataset alongside an international one enriches the analysis, providing a broader understanding of demand prediction in diverse markets and enhancing the robustness and applicability of predictive models across different regions. Additionally, considering the different types of food in each dataset, this comparison allows for more tailored demand prediction strategies that account for the specific characteristics and consumer behaviors associated with online grocery shopping versus food delivery.

4. METHODOLOGY

This section outlines the methodology for developing and evaluating machine learning algorithms for demand prediction, involving key steps: data collection, data preprocessing, feature engineering, model selection, model training, and evaluation. Data collection and feature engineering were explained in the previous section. Here, we delve into data preprocessing and explain all the models in detail.

4.1 Data Preprocessing

In order to use the data in the machine learning models, preprocessing the data and transforming categorical features into a numerical feature is an essential step. This transformation can be accomplished in several ways, each of which is appropriate for a particular kind of categorical data. The two most well-known techniques are one hot encoding and label encoding. Binary vectors are used to represent categorical variables in one-hot encoding. Given the presence of a high number of unique values in our categorical variables, we preferred using label encoding instead of one-hot encoding. One-hot encoding, while effective for small to moderate numbers of categories, becomes impractical with a large number of unique values as it results in a very high-dimensional feature space (Hinton, 1993). This can lead to a significant increase in computational complexity and memory usage, potentially causing performance issues. Additionally, models can struggle with overfitting due to the sparsity of the resulting dataset (Huber et al., 2017). In contrast, label encoding assigns a unique integer to each category, maintaining a compact representation without increasing the dimensionality of the dataset. This approach is computationally efficient and helps avoid the pitfalls associated with high-dimensionality in one-hot encoding. Thus, label encoding is a practical and efficient choice for transforming categorical features with numerous unique values like our dataset. Preparing the data for machine learning algorithms that need numerical input is the primary goal. In our dataset,

we transformed the categorical features; category, cuisine, center type into numerical values using label encoding.

4.2 Models

As baseline models, we started with several well-known machine learning models: Linear Regression, Decision Tree Regressor, Random Forest Regressor, GBR, and XGBoost. Moreover, a time series model called NARX was applied to our dataset.

When dealing with perishable goods, efficient demand prediction is critical due to the products' short shelf life and the need for quick decision-making to minimize waste and optimize inventory. In this context, the computational efficiency and speed of machine learning models become paramount. For our datasets, the trained models were efficient and fast. Linear Regression is an excellent choice for scenarios requiring minimal computational resources and quick predictions. Its simplicity ensures fast training and inference times, making it suitable for real-time applications where rapid response is essential. However, its capacity to capture complex, non-linear relationships in the data is limited, potentially impacting prediction accuracy for more intricate datasets.

Decision Tree Regressor offers a balance between speed and the ability to handle non-linear relationships. It is moderately resource-intensive and provides interpretable results, which can be beneficial for understanding the factors driving demand. Nonetheless, decision trees are prone to overfitting, particularly with smaller datasets, which can reduce their generalizability.

Random Forest Regressor addresses some of the overfitting issues seen with single decision trees by averaging the results of multiple trees. This ensemble method improves predictive accuracy and robustness at the cost of increased computational resources. The training process, involving the construction of numerous trees, can be time-consuming, but prediction times remain relatively fast, making it a viable option for many applications. GBR and XGBoost represent more advanced techniques that build models in a sequential manner to correct errors made by previous models. These methods often yield higher predictive accuracy and can model complex relationships in the data effectively.

However, they are significantly more resource-intensive and slower to train compared to simpler models like linear regression or decision trees. XGBoost is optimized for speed and performance, utilizing techniques like parallel processing and tree pruning, making it more efficient than traditional gradient boosting methods.

The NARX model, specifically designed for time series data, is highly suitable for predicting demand for perishable goods. It can capture both the temporal dynamics and external factors affecting demand, offering a comprehensive approach to forecasting. The complexity of the NARX model requires substantial computational resources and longer training times, but its ability to model intricate temporal patterns can lead to superior predictive performance.

In addition to these models, other approaches such as LSTM networks could be considered. LSTM networks are a type of recurrent neural network particularly adept at handling sequential data and long-term dependencies, which is beneficial for time series forecasting. However, they are computationally demanding and require significant training time.

Ultimately, the choice of model depends on the specific requirements of the application, including the complexity of the data, the need for interpretability, and the available computational resources. For perishable goods, a combination of simpler models for real-time predictions and more complex models for periodic, in-depth forecasting could provide an optimal balance between speed and accuracy. For this thesis, we have started off with more simple models like linear regression and progressively we moved on to the more complex ones. The models in this thesis were selected with an extensive literature review and analysis of the most popular prediction models. Moreover, the NARX model was chosen for being an innovative and popular model for time series forecasting. This section explains each model in detail.

4.2.1 Linear Regression (LR)

LR is a foundational algorithm in machine learning, frequently used for predictive modeling and data analysis. It is based on the concept of fitting a linear relationship between

input variables (features) and an output variable (target). The primary goal is to find the linear function that best predicts the output variable by minimizing the sum of the squared differences between the observed values and the values predicted by the linear function. It is a statistical method for modelling the relationship between a dependent variable and one or more independent variables. With LR, the model predicts the dependent variable using a single independent variable. One of the main advantages of LR is its simplicity and interpretability. The model assumes a linear relationship between the independent variables and the dependent variable, making it straightforward to understand and implement. It provides clear insights into how each input variable affects the output, which can be valuable for understanding underlying patterns in the data (Ke et al., 2017). However, LR also has limitations. It assumes that there is a linear relationship between the variables, which may not always hold true in real-world datasets. Additionally, it is sensitive to outliers in data, which can significantly affect the model's performance (Lutoslawski et al., 2021). Despite these limitations, linear regression remains a widely used technique, especially as a baseline model for more complex algorithms.

4.2.2 Decision Tree Regressor (DTR)

A DTR is a machine learning model used for predictive modeling tasks. It is a powerful predictive modeling technique and has garnered significant attention in the realm of machine learning and data analysis. In their paper, "Decision Trees for Decision Making", Breiman elucidates the fundamental principles underlying decision tree construction and its applicability in regression tasks (Breiman et al., 2017). Decision trees partition the feature space into distinct regions, each associated with a constant value prediction, enabling interpretable and intuitive modeling of complex relationships between input variables and continuous target outcomes. Furthermore, the flexibility of decision trees allows for non-linear decision boundaries and automatic feature selection, making them adept at capturing intricate patterns within data (Breiman et al., 2017). Quinlan's seminal work on regression trees (1992) emphasizes the interpretability and computational efficiency of decision tree regression, particularly in domains where understanding the decision-making process is as crucial as predictive accuracy (Quinlan, 1992). Through recursive binary splitting, decision trees iteratively divide the feature space based on optimal splitting criteria, such as minimizing mean squared error, thus constructing a hierarchical structure

of decision nodes that efficiently approximate the underlying regression function. Additionally, decision tree regression exhibits resilience to outliers and missing values, providing robustness in real-world datasets (Quinlan, 1992). With advancements in ensemble techniques like Random Forests and Gradient Boosting, decision tree regression continues to be a cornerstone in predictive modeling, offering a blend of simplicity, interpretability, and predictive performance that caters to diverse analytical needs.

4.2.3 Random Forest Regressor (RFR)

RFR, an ensemble technique derived from decision trees, has emerged as a robust and widely used tool for regression tasks in machine learning. Breiman's seminal work on Random Forests (2001) introduced the concept of aggregating multiple decision trees trained on random subsets of the data and features to achieve higher predictive performance and resilience to overfitting (Breiman, 2001). By averaging the predictions of individual trees or taking a mode for classification tasks, RFR effectively mitigates the variance inherent in single decision trees while preserving their interpretability and flexibility. Furthermore, the inherent randomness in feature selection and data sampling during tree construction enhances model generalization and reduces the risk of model bias. Subsequent studies have underscored the superior predictive accuracy and robustness of RFR across diverse domains, making it a favored choice for regression problems in fields ranging from finance to healthcare. With its ability to handle large-scale datasets, nonlinear relationships, and high-dimensional feature spaces, RFR continues to be a cornerstone in predictive modeling, offering a potent blend of accuracy, interpretability, and scalability (Liaw & Wiener, 2002).

4.2.4 XGBoost

XGBoost, a cutting-edge gradient boosting algorithm, has revolutionized the landscape of machine learning with its exceptional predictive accuracy and computational efficiency. Developed by Chen and Guestrin (2016), XGBoost employs an innovative gradient boosting framework that iteratively builds an ensemble of weak learners, typically decision trees, to sequentially minimize a differentiable loss function (Chen & Guestrin,

2016). By leveraging a combination of gradient descent optimization techniques, regularization strategies, and tree pruning algorithms, XGBoost achieves superior predictive performance while mitigating overfitting and enhancing model interpretability. The scikit-learn library was used for this model. Through the *XGBClassifier* and *XGBRegressor* classes, Scikit-learn offers a wrapper for XGBoost, enabling users to apply XGBoost (PedregosaFabian et al., 2011).

4.2.5 Gradient Boosting Regressor (GBR)

GBR, a prominent ensemble learning technique, has gained widespread acclaim for its remarkable predictive accuracy and versatility in handling complex regression tasks. Initially proposed by Friedman (2001), GBR operates by iteratively improving upon the shortcomings of preceding weak learners, thereby incrementally minimizing a specified loss function. This iterative optimization process involves fitting new models to the residuals of the previous predictions, thereby progressively refining the predictive performance of the ensemble (Friedman, 2001). Through the fusion of individual weak learners into a robust and adaptive ensemble, GBR excels at capturing intricate nonlinear relationships and interactions within the data. Moreover, the algorithm's inherent capacity for handling heterogeneous data types, missing values, and outliers contributes to its resilience and generalization capability. Notably, the XGBoost library, an optimized implementation of gradient boosting, has further popularized the adoption of GBR due to its computational efficiency and scalability (Chen & Guestrin, 2016). With its ability to deliver state-of-the-art performance across diverse domains, GBR continues to be a cornerstone in predictive modeling, offering a potent blend of accuracy, interpretability, and computational efficiency.

4.2.6 Multi-layer Perceptron (MLP) Regressor

MLPRegressor is a powerful tool in the realm of machine learning, designed for performing regression tasks using a multi-layer perceptron (MLP) architecture. As part of the scikit-learn library, MLPRegressor utilizes neural network principles to model complex relationships between input and output variables. It features a network of interconnected

neurons organized in layers, including input, hidden, and output layers. Each neuron applies a linear transformation to its input and passes the result through a non-linear activation function, enabling the network to capture intricate patterns in the data (PedregosaFabian et al., 2011). MLPRegressor supports various hyperparameters, such as the number of hidden layers and neurons, learning rate, and activation function, allowing users to tailor the model to specific datasets and regression problems. It employs backpropagation for training, which involves adjusting the weights of the network to minimize the mean squared error between predicted and actual values. While MLPRegressor is highly flexible and capable of achieving high accuracy, it requires careful tuning and can be computationally intensive, especially for large datasets and deep network (Hinton, 1993). The algorithm's performance also depends on the quality and quantity of the input data, necessitating thorough preprocessing and feature engineering (Goodfellow et al., 2016).

4.2.7 Light Gradient Boosting Machine (LightGBM)

LightGBM is a sophisticated and highly efficient framework for gradient boosting, developed by Microsoft. It is designed to handle large datasets with high performance and scalability, making it a popular choice for machine learning tasks, particularly in competitions and industrial applications (Ke et al., 2017). LightGBM employs a leaf-wise tree growth strategy as opposed to the level-wise approach used by many other gradient boosting algorithms. This allows it to reduce more loss and grow deeper trees, which often results in better accuracy. Additionally, it supports histogram-based algorithms, which significantly speed up the training process by binning continuous features into discrete bins, thus reducing memory consumption and computational complexity (Ke et al., 2017).

4.2.8 Nonlinear Autoregressive model with Exogenous Inputs (NARX)

The NARX represents a sophisticated approach to time series forecasting, integrating both the autoregressive nature of time series data and the influence of external variables. As detailed by Chen and Billings (1989), the NARX model captures the nonlinear dependencies in time series data by regressing the current value of the series on its past values and on past values of the exogenous inputs (Chen & Billings, 1989). This dual consideration of endogenous and exogenous factors allows NARX models to effectively model complex systems where external factors significantly impact the behavior of the

time series. Utilizing techniques such as neural networks for nonlinear mapping, NARX models excel in capturing intricate patterns and dynamic relationships that linear models may overlook. The flexibility of NARX models to accommodate various forms of nonlinearity and their robustness in handling different types of input signals make them particularly valuable in fields like control systems, economics, and environmental modeling (Breiman, 2001). By leveraging both historical data and relevant external inputs, NARX models provide a comprehensive framework for accurate and insightful time series forecasting, thereby enhancing decision-making processes across numerous applications. It differs from traditional regression models in several significant ways, making it particularly advantageous in specific scenarios. Unlike traditional regression models, which often assume independent observations and are used for static data, NARX is specifically designed for time series data where the value at any time point depends on past values and possibly past values of exogenous (external) inputs. NARX incorporates autoregressive terms, utilizing past values of the dependent variable to predict future values, thus capturing temporal dependencies more effectively than traditional models. Furthermore, NARX is adept at modeling nonlinear relationships within the data through the use of neural networks or other nonlinear functions, unlike many traditional regression models that are linear or assume specific forms of nonlinearity. Additionally, NARX explicitly incorporates exogenous variables that may influence the dependent variable, providing a more flexible framework for scenarios where outside factors are significant. The use of NARX models is particularly beneficial for capturing temporal dynamics, incorporating external influences, and modeling nonlinear relationships, leading to better performance in fields such as economics, environmental science, and control systems. This flexibility and robustness make NARX models essential for accurate and insightful predictions where traditional regression models might fall short (Chen & Billings, 1989).

4.3 Model Selection

For demand prediction, we meticulously selected a range of machine learning models to leverage their diverse strengths and capabilities. Each model was chosen for its unique advantages in capturing various aspects of demand prediction, ensuring a robust and accurate forecasting system. LR was selected for its simplicity and interpretability. It

assumes a linear relationship between the input features and the target variable, which, despite being a basic assumption, often provides a solid foundation for understanding the fundamental relationships in the data. Its computational efficiency makes it ideal for small to moderately large datasets, serving as an excellent baseline model against which the performance of more complex models can be measured.

The DTR was included due to its ability to capture non-linear relationships and its flexibility in handling both numerical and categorical data without requiring extensive preprocessing. Its structure allows for straightforward interpretation through visualization, making it easier to understand the decision-making process. This model's capability to handle complex and varied data types is particularly valuable in demand prediction scenarios where diverse factors influence the target variable.

To enhance prediction accuracy and robustness, we employed the RFR. This approach mitigates the risk of overfitting, a common issue with individual decision trees, and provides insights into feature importance, helping to identify the most influential factors in predicting demand. By leveraging the power of ensemble learning, Random Forest Regressor offers a balanced trade-off between bias and variance, leading to more reliable predictions.

GBR and XGBoost were selected for their powerful ensemble techniques that build trees sequentially, with each new tree correcting errors made by its predecessors. This sequential learning process enhances predictive accuracy and captures complex patterns and interactions within the data. GBR's flexibility allows for fine-tuning through various hyperparameters, enabling the model to adapt to different datasets and prediction tasks effectively. XGBoost, an optimized implementation of gradient boosting, is renowned for its speed and performance, scalability, and advanced features such as handling missing data and tree pruning. These capabilities make XGBoost particularly suitable for large-scale demand prediction tasks where efficiency and robustness are critical.

The inclusion of the NARX model addresses the temporal aspect of demand prediction. NARX is specifically designed for time series data, capable of modeling dynamic temporal dependencies and incorporating external variables that influence the target variable.

This model's ability to capture non-linear relationships and temporal dependencies makes it ideal for complex demand prediction tasks where past values and external factors significantly impact future demand. By integrating these diverse models, we aim to create a comprehensive and accurate demand forecasting system that leverages the strengths of each approach, accommodates both linear and non-linear relationships, handles large datasets efficiently, and incorporates temporal dynamics and external influences to provide robust and reliable predictions.

5. EVALUATION METRICS AND RESULTS

In this section, we compared the performance of various machine learning models on a prediction task to evaluate the model performance and guide improvements. The results will be given in two sections. The first section will give the results for Online Market Dataset and the second section will give the results for Kaggle dataset.

5.1 Online Market Dataset Results

The results for Online Market Dataset are in Table 5.1.1.

TABLE 5.1.1: Comparison of Metrics from All the Models on Online Market Test Dataset

Models	R^2
Linear Regression	0,31
Decision Tree Regressor	0,62
Random Forest Regressor	0,63
Gradient Boosting Regressor	0,50
XGBoost	0,57
MLPRegressor	0,45
LightGBM	0,55

The table presents the R-squared (R^2) values of different regression models, indicating their performance in explaining the variance in the data. R^2 values range from 0 to 1, where a higher value signifies better explanatory power.

Linear Regression model has an R^2 value of 0,31, suggesting that it explains 31% of the variance in the data. This is relatively low, indicating that Linear Regression might not be capturing the complexity of the data well.

Decision Tree Regressor, with an R^2 value of 0,62, this model performs significantly better than Linear Regression. It explains 62% of the variance, indicating a better fit to the data.

Random Forest Regressor model has an R^2 value of 0,63, slightly higher than the Decision Tree Regressor. The improvement suggests that the ensemble method of combining multiple decision trees is beneficial.

Gradient Boosting Regressor model has an R^2 value of 0,50, lower than the Decision Tree and Random Forest Regressors but higher than Linear Regression. It indicates that while Gradient Boosting is effective, it may not be the best model for this particular dataset.

XGBoost model has an R^2 value of 0,57, which is a competitive performance, showing that it explains 57% of the variance. XGBoost is known for its high performance and efficiency in many datasets.

MLPRegressor has an R^2 value of 0,45, indicating moderate performance. It shows that the neural network model captures 45% of the variance in the data.

LightGBM model has an R^2 value of 0,55, which is close to the performance of XGBoost. It suggests that LightGBM is also effective in capturing the variance in the data.

In summary, the RFR shows the highest R^2 value (0,63), indicating the best performance among the models listed. The DTR also performs well, followed closely by XGBoost and LightGBM. LR shows the lowest R^2 value, indicating that it might not be the best choice for this dataset.

As stated in the data analysis sectioni most of the dataset is from the categories Fruits, Vegetables and Organic Products. These categories are perishable goods; therefore, we believed a separate model just for the perishable goods would have higher accuracy. For this reason, we divided the dataset and created a smaller dataset with the categories Fruits, Vegetables and Organic Products. This smaller dataset has 2.240.284 records. All the seven models were trained for this dataset as well. The results are in Table 5.1.2.

TABLE 5.1.2: Comparison of Metrics from All the Models on Online Market Test Dataset For Perishable Goods

Models	R^2
Linear Regression	0,41
Decision Tree Regressor	0,70
Random Forest Regressor	0,79
Gradient Boosting Regressor	0,61
XGBoost	0,75
MLPRegressor	0,38
LightGBM	0,76

After creating the new dataset based on perishable food categories and retraining the models, significant improvements were observed in the performance of most models. The LR model showed a modest increase in R^2 from 0,31 to 0,41, indicating a better fit. The DTR also improved, with its R^2 increasing from 0,62 to 0,70. Notably, the RFR and XGBoost models demonstrated substantial enhancements, with their R^2 values rising from 0,63 to 0,79 and 0,57 to 0,75, respectively. The GBR and LightGBM models also experienced notable gains, with the GBR increasing from 0,50 to 0,61 and LightGBM from 0,55 to 0,76. Conversely, the MLPRegressor's performance declined, with its R^2 dropping from 0,45 to 0,38. Overall, the removal of certain categories appears to have positively impacted the predictive capabilities of most models, particularly the ensemble methods like RFR, XGBoost, and LightGBM, which showed the most significant improvements. This result indicates that modeling perishable goods alone increases the accuracy. It is also better to handle these products in a separate model because the risk of spoilage would be minimum with higher accuracy thus reducing the waste. On the other hand, with the removal of this categories the dataset count is significantly dropped. Certain products don’t have enough records to train a model so if we separate the perishable goods we might need to have more records on other categories.

For this reason, we believed hierarchical learning could be useful. Hierarchical learning is an advanced methodology in machine learning that structures learning models into multiple layers to capture various levels of abstraction and complexity in data. This approach mirrors the hierarchical nature of human learning, where simple concepts are understood understood first and then combined to understand more complex concepts. Hierarchical

learning aims to improve model performance and reduce computational complexity by leveraging multiple levels of classifiers, each focusing on different aspects of the data (Zhang & Zhang, 2006).

In hierarchical learning, a general model is first trained to capture broad patterns across the entire dataset. This general model helps in understanding the overall structure and general trends in the data. Following this, specialized models are trained on subsets of the data that exhibit unique characteristics, such as perishable goods in the context of your dataset. These specialized models are designed to capture the nuances and specific patterns within these subsets that the general model might miss (Zhang & Zhang, 2006).

The hierarchical approach offers several benefits. It allows for more accurate predictions by ensuring that the unique attributes of different data subsets are effectively captured. For instance, in a dataset containing both perishable and non-perishable goods, a single model might not adequately capture the distinct demand patterns of perishable items. By training a specialized model specifically for perishable goods, hierarchical learning ensures that these unique patterns are well understood and accurately predicted.

In the provided methodology, hierarchical learning is implemented by first loading the dataset and labeling perishable products with a new column, is_perishable. The data is then split into features and target variables, and further divided into training and testing sets. A general model is trained on the entire dataset using the RFR algorithm because for this dataset, RFR had the highest R^2 values. Predictions from this general model are then added as a new feature to both the training and testing datasets.

Subsequently, the data is filtered to create a subset containing only perishable items, and specialized models—such as RFR, XGBRegressor, and LGBMRegressor—are trained on this subset. We chose these 3 models because they have the highest R^2 value for the subset dataset with only perishable products. A hierarchical prediction function is defined to determine which model to use for making predictions based on whether a product is perishable. This function ensures that perishable products are predicted using the specialized model, while non-perishable products are predicted using the general model.

The performance of the combined hierarchical model is evaluated using the R^2 metric. By combining the general model and specialized models, hierarchical learning effectively captures both broad and specific patterns in the data, leading to improved predictive performance and robustness. This approach ensures that the unique characteristics of different product categories are comprehensively modeled, resulting in more accurate and reliable predictions.

The hierarchical model yielded an R^2 value of 0,51, reflecting its capacity to balance the complexity and specificity of different product categories. While this performance is a marked improvement over the LR model and is competitive with other models like GBR and MLPRegressor, it is still lower than the performance of specialized models like the RGR (0,63) and XGBoost (0,57) on the overall dataset. However, it is essential to highlight that the hierarchical approach provides a more nuanced prediction mechanism, effectively capturing the distinct demand patterns of both perishable and non-perishable goods. This balance between broad and specialized predictive capabilities ensures more reliable forecasting, particularly valuable in contexts where accurate prediction of perishable goods' demand is crucial for reducing waste and optimizing supply chain operations. Overall, the hierarchical learning methodology demonstrates a promising avenue for enhancing model performance, particularly in datasets characterized by diverse and distinct subcategories.

5.2 Food Demand Forecasting Dataset Results

The results for Food Demand Forecasting datasets are in Table 5.2.1.

TABLE 5.2.1: Comparison of Metrics from All the Models on the Test Dataset

Models	RMSE	R^2	MSE	MAE
Linear Regression	346,4	0,25	119.990,55	195,82
Decision Tree Regressor	226,4	0,68	51.254,93	92,53
Random Forest Regressor	150,38	0,86	22.614,96	68,94
Gradient Boosting Regressor	243,42	0,63	59.250,9	124,20
XGBoost	154,8	0,84	23.962,8	80,72
MLPRegressor	194,91	0,75	37.992,34	104,97
LightGBM	138,45	0,87	19.169,66	75,51

By looking at the results, we concluded that of all the models, LR performed the least well; its RMSE was 346,4, indicating a high prediction error. The R^2 value of 0,25 signifies that the model explains only 25% of the variance in the data, which is quite low. The MSE is extremely high at 119.990,55, indicating significant discrepancies between the predicted and actual values. The MAE of 195,82 further emphasizes the model's limited accuracy, with large average errors in its predictions. According to these findings, LR had the highest error rates and explained the least variance in the dataset.

The DTR shows improved performance over LR, with an RMSE of 226,4. The R^2 value of 0,68 suggests that the model accounts for 68% of the variance, indicating a moderate fit. The MSE is 51.254,93, and the MAE is 92,53, reflecting that the model has a better prediction accuracy than LR. It demonstrated a noteworthy enhancement over LR but still has room for improvement. This model accounted for a greater percentage of the data volatility and significantly decreased the error metrics.

Performance was further improved with the RFR, with an RMSE of 150,38. The R^2 value is high at 0,86, meaning it explains 86% of the variance in the data, indicating a strong fit. The MSE is considerably lower at 22.614,96, and the MAE is 68,94, demonstrating that the model is quite effective at making accurate predictions with lower average errors. This model greatly reduced error rates and increased the explained variance over the DTR.

The GBR provided intermediate results, with an RMSE of 243,42, which is higher than Random Forest but still reasonable. The R^2 value is 0,63, explaining 63% of the variance, indicating a moderate fit. The MSE is 59.250,9, and the MAE is 124,20, suggesting that while the model is accurate, it is less so compared to the RFR.

The XGBoost model shows good performance with an RMSE of 154,8 and an R^2 value of 0,84, indicating that it explains 84% of the variance. The MSE is 23.962,8, and the MAE is 80,72, showing that the model is effective and accurate in its predictions, with relatively low error metrics.

The MLPRegressor has an RMSE of 194,91, which is moderate, and an R^2 value of 0,75, explaining 75% of the variance. The MSE is 37.992,34, and the MAE is 104,97, suggesting that while the model is reasonably accurate, it is less effective compared to XGBoost and RFR.

LightGBM model exhibits the best performance among all models with an RMSE of 138,45. The R^2 value is the highest at 0,87, explaining 87% of the variance, indicating a very strong fit. The MSE is the lowest at 19.169,66, and the MAE is 75,51, reflecting that the model is highly accurate with the smallest average errors in its predictions.

In summary, LightGBM demonstrates the highest accuracy and best overall performance, followed by the RFR and XGBoost among the models tried in this thesis. The LR model performs the poorest, indicating the need for more complex models to achieve better prediction accuracy.

Moreover, we assessed the performance of a NARX model with the train dataset on the prediction challenge in addition to the conventional machine learning models. The results are in the tables 5.2.2 and figures 5.2.1. The provided plots show the last 100 timesteps of predicted values and true values.

The evaluation of the NARX model on the training dataset provides insightful metrics that demonstrate its performance. The RMSE for the NARX model is 237,74, indicating a reasonable level of prediction error within the training data. The R^2 is 0,61, suggesting that the model can explain 61% of the variance in the training dataset. This value indicates a moderate level of fit, where the model captures a significant portion of the data's variability, yet there is still room for improvement.

The MSE is reported as 56.518,06, reflecting the average squared differences between the predicted and true values. A lower MSE is desirable, and while the NARX model's MSE shows that it is capable of learning from the training data, the value also suggests that some predictions are notably off from the actual values. Additionally, the MAE for the model is 107,21, which indicates that, on average, the model's predictions deviate

from the true values by approximately 107 units. This metric provides a more interpretable measure of the average magnitude of errors.

The accompanying plot of the last 100 timesteps further illustrates the model's performance. The predicted values are shown in one color, and the true values in another, allowing for a visual comparison. The model generally follows the true values, capturing the overall trend and significant peaks and valleys. However, there are still noticeable deviations, particularly around the higher peaks, where the model's predictions either overshoot or undershoot the actual values.

In summary, the NARX model demonstrates a decent performance on the training dataset with an RMSE of 237,74, an R^2 of 0,61, an MSE of 56.518.06, and an MAE of 107.21. These metrics indicate that the model has a good grasp of the data's underlying patterns but also highlight areas where the model can be improved to reduce prediction errors and increase accuracy.

TABLE 5.2.2: Comparison of Metrics from NARX Model on the Train Dataset

Models	RMSE	R^2	MSE	MAE
NARX	237,74	0,61	56.518,06	107,21

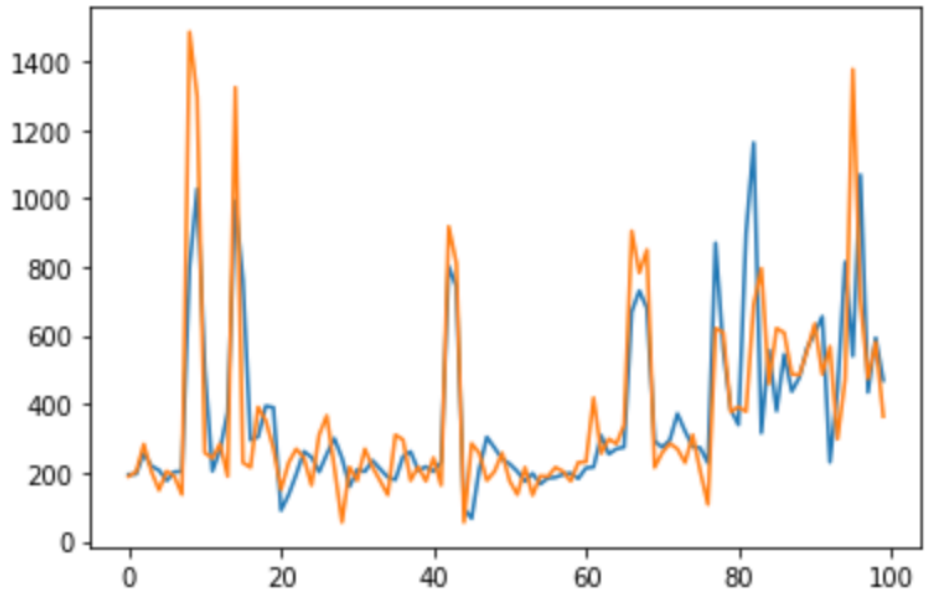

FIGURE 5.2.1 – Last 100 timesteps with True and Predicted Values on Train Dataset

For the test dataset the results are in the figure 5.2.2 and Table 5.2.3. In the figure, predicted values generally follow the true values. This indicates that the model captures the overall trend and patterns in the data. At certain points, there are noticeable deviations between the predicted and true values. For example, around the 20th and 80th timesteps, the model prediction diverges significantly from the true values. The model is able to capture the peaks and valleys in the true values to a reasonable extent, although not perfectly.

The RMSE of 280,7 indicates that there are substantial errors in the model's predictions, highlighting a significant deviation between the predicted and true values. The R^2 value of 0,58 suggests that the model accounts for 58% of the variance observed in the data, reflecting a moderate level of predictive accuracy. The MSE stands at 78792.58, further emphasizing the presence of large prediction errors. Additionally, the MAE is 117.23, which means that, on average, the model's predictions deviate from the true values by 117,23 units. These metrics collectively indicate that while the model can capture general trends in the data, there is considerable room for improvement to enhance its accuracy and reduce prediction errors.

TABLE 5.2.3: Comparison of Metrics from NARX Model on the Test Dataset

Models	RMSE	R^2	MSE	MAE
NARX	280,7	0,58	78792,58	117,23

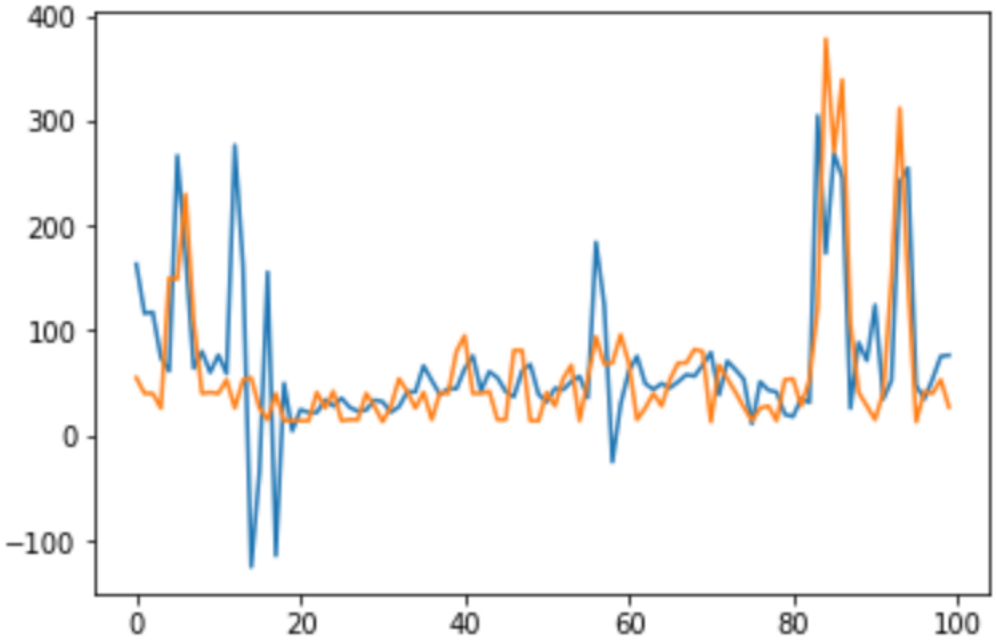

FIGURE 5.2.2 – Last 100 timesteps with True and Predicted Values on Test Dataset

Furthermore, the model was trained again with other features. In developing a new NARX model with an enhanced set of features, the aim was to improve the model's predictive accuracy and overall performance. The initial model, which utilized features such as category, cuisine, center type, checkout price, and week provided a foundational understanding of the data but fell short in capturing the complexity and variance within the dataset. To address this, a new set of features was introduced, including week, center_id, meal_id, checkout price, base price, emailer for promotion, and homepage featured. These features were chosen for their potential to provide a more comprehensive view of the factors influencing the number of orders. For instance, "center_id" and "meal_id" introduced specific identifiers that allowed the model to capture unique patterns related to different centers and meals. The addition of "base_price" alongside "checkout_price" offered deeper insights into pricing dynamics, while the inclusion of promotional features such as

"emailer_for_promotion" and "homepage_featured" helped in understanding the impact of marketing activities. By incorporating these relevant and diverse features, the new NARX model was better equipped to capture the underlying trends and patterns in the data, resulting in improved performance metrics and a more robust predictive capability.

For the train dataset, the NARX model achieved a Root Mean Square Error (RMSE) of 217,54, indicating the average magnitude of the errors between predicted and actual values. The R^2 value of 0,67 suggests that 67% of the variance in the dataset is explained by the model, indicating a decent fit. The Mean Square Error (MSE) is 47.319,51, providing a measure of the average of the squares of the errors. Lastly, the Mean Absolute Error (MAE) is 101,26, representing the average absolute difference between predicted and actual values.

TABLE 5.2.4: Comparison of Metrics from NARX Model on the Train Dataset with New Features

Models	RMSE	R^2	MSE	MAE
NARX	217,54	0,67	47.319,51	101,26

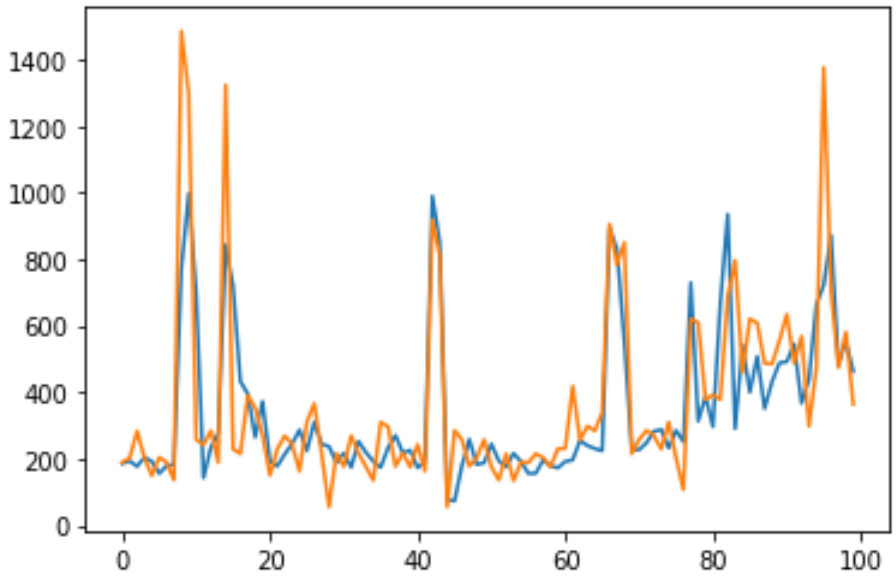

FIGURE 5.2.3 – Last 100 timesteps with True and Predicted Values on Train Dataset with New Features

On the test dataset, the model's performance showed a slight decrease compared to the

test with the train dataset. The RMSE increased to 257,48, and the R^2 value slightly decreased to 0,65, showing that the model explains 65% of the variance in the test data. The MSE is 66.298,52, indicating a higher average squared error compared to the training set. The MAE for the test dataset is 110,64, reflecting a slight increase in the average absolute error compared to the training data.

TABLE 5.2.5: Comparison of Metrics from NARX Model on the Test Dataset with New Features

Models	RMSE	R^2	MSE	MAE
NARX	257,48	0,65	66.298,52	110,64

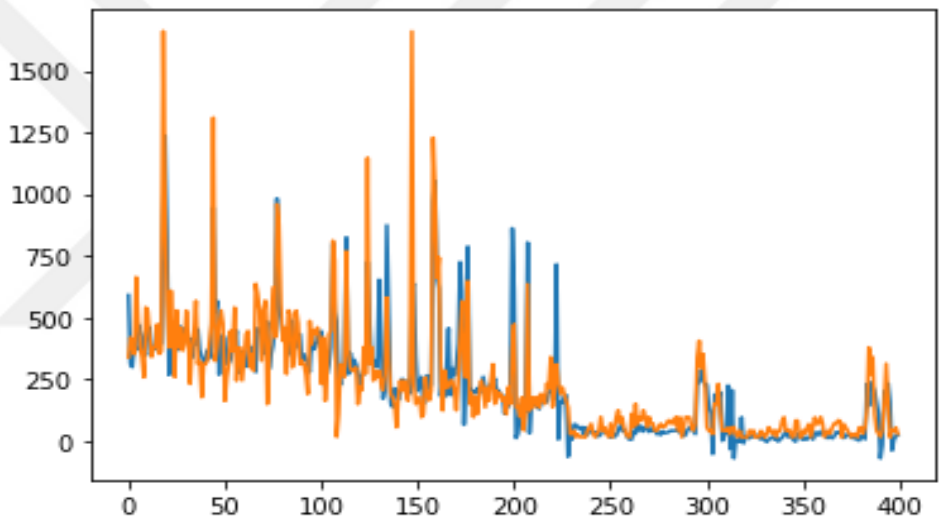

FIGURE 5.2.4 – Last 100 timesteps with True and Predicted Values on Test Dataset with New Features

Figures 5.2.3 and 5.2.4 display the last 100 timesteps of true and predicted values for the train and test datasets, respectively. These visual representations help in assessing the model's predictive performance over time. In both figures, the predicted values (represented by the lines) closely follow the true values, indicating that the model is capturing the underlying patterns in the data, although with some discrepancies, especially during peaks.

Overall, on the second model, NARX model demonstrates reasonable performance in predicting the time series data, with slightly better results on the training set compared to

the test set. The metrics and visual analysis suggest that the model is effective but may benefit from further tuning or additional features to improve its predictive accuracy.

In addition, comparing the first and second NARX model and their results for the test dataset reveals significant improvements in the current NARX model with new features over the previous model. The RMSE has decreased from 280,7 to 257,48, indicating a reduction in the overall prediction error. Additionally, the R^2 value has improved from 0,58 to 0,65, suggesting that the current model explains a greater proportion of the variance in the data, thereby offering a better fit.

Moreover, the MSE has decreased from 78.792,58 to 66.298,52, reflecting a lower average of the squared errors and thus better predictive performance. The MAE has also been reduced from 117,23 to 110,64, indicating that the average absolute deviation of the predictions from the true values has decreased, signifying more accurate predictions on average.

Visual analysis of the figures shows that in both models, the predicted values generally follow the true values, capturing the overall trends and patterns in the data, including peaks and valleys. However, the current model demonstrates closer alignment with the true values, with less pronounced deviations, particularly around specific timesteps.

In summary, the current NARX model with new features exhibits notable improvements over the previous version, as evidenced by the reduced RMSE, MSE, and MAE, and the higher R^2 value. While the model effectively captures the general trends and patterns, there is still room for further refinement to enhance accuracy and reduce prediction errors. Moreover, if we were to compare the second NARX model with the benchmark models, we saw that this time the NARX model was better than GBR and LR and it was very close to the DTR model in terms of R^2 value. This proves that as future work more in-depth study on feature engineering and different combinations of features better models with higher accuracies could be achieved.

5.3 Discussion

The evaluation of various machine learning models across the Online Market Dataset and the Kaggle Food Demand Forecasting Dataset revealed several critical insights into demand forecasting and model performance. Notably, the RFR consistently demonstrated robust performance across both datasets. On the Online Market Dataset, the RFR achieved the highest R^2 value of 0,63, indicating a strong ability to explain the variance in this dataset, which included a diverse range of products. Similarly, on the Kaggle dataset, the RFR's performance was exceptional, with an R^2 value of 0,86 and an RMSE of 150,38, underscoring its robustness in different contexts. The consistent performance of RFR highlights its versatility and efficacy in handling diverse and complex datasets, making it a reliable choice for demand forecasting tasks.

In contrast, the LR model consistently underperformed, with R^2 values of 0,31 on the Online Market Dataset and 0,25 on the Kaggle dataset. These results suggest that simple linear models fail to capture the intricate and non-linear patterns present in these datasets. The high prediction errors and low explanatory power of the LR model underline the necessity for more sophisticated modeling techniques in demand forecasting, where capturing complex interactions and variabilities is crucial for accurate predictions.

The separation of perishable goods from the Online Market Dataset into a specialized subset revealed significant improvements in model performance. For instance, the RFR and XGBoost models showed substantial enhancements, with R^2 values increasing to 0,79 and 0,75, respectively. This finding highlights the importance of tailored modeling approaches for different product categories. Perishable goods, with their unique demand patterns influenced by factors like seasonality and spoilage risk, benefit from specialized models that can better capture these dynamics. The improved accuracy for perishable goods suggests that specialized models are crucial for optimizing inventory management and reducing waste in the supply chain.

The hierarchical learning approach, which involved creating a general model for the entire dataset and specialized models for subsets, provided a balanced prediction mechanism.Although the hierarchical model's R^2 value of 0,51 did not surpass the specialized models like RFR and LightGBM, it effectively captured the unique characteristics of both perishable and non-perishable goods. This approach demonstrates potential for improving predictive accuracy in datasets with heterogeneous categories by combining the broad understanding of a general model with the detailed focus of specialized models. The hierarchical learning method's nuanced prediction mechanism is particularly valuable in contexts where different product categories exhibit distinct demand patterns, as it ensures more accurate forecasting across the entire dataset.

The Kaggle dataset's evaluation further reinforced the superior performance of advanced models like LightGBM, which achieved an R^2 of 0,87 and the lowest RMSE of 138,45. LightGBM's ability to handle large-scale data efficiently and capture complex interactions made it particularly effective in this context. The inclusion of additional features in the Kaggle dataset significantly enhanced the performance of models like NARX, underscoring the critical role of feature engineering. By incorporating relevant features that capture important aspects of the data, models can achieve higher accuracy and provide more reliable predictions.

The comparative analysis across the two datasets emphasizes the importance of understanding the specific characteristics of the data and the forecasting problem at hand. While models like RFR and LightGBM consistently showed high performance, the variability in results across different datasets highlights the necessity for comprehensive evaluation processes.

6. CONCLUSION

Demand forecasting is an important topic for effective supply chain management and business planning. It contributes to companies and organizations, providing valuable insights about future demand. Working on demand forecasting gives firms and individuals a way to optimize inventory levels, helps them make efficient and well-planned financial decisions, and reduces possible risks to the minimum. Plus, it enhances the customer experience since the companies would be able to provide the demanded product in the correct quantities at the correct time to meet customer's wishes.

Moreover, since forecasting demand helps manufacturers to coordinate their production and storage, it ensures that the amount of perishable goods would meet the real demand. It will reduce overproduction, spoilage and waste. In addition, demand forecasting supports sustainable practices since it lowers the wasted products and resources used. Not only does it contribute to companies and humans, but it also supports sustainability goals benefiting the environment and the world by reducing waste, conserving resources, and minimizing the ecological footprint.

This thesis explores a variety of techniques, algorithms, and tactics within the complex field of demand forecasting. The main goal of this thesis is to increase our understanding of demand prediction by thoroughly reviewing and analyzing the most recent forecasting algorithms and approaches. Two datasets were used in this study. Firstly, we have worked on an Online Market Dataset. Moreover, we have used a dataset from Kaggle. Each dataset contains food-related data.

The models used in this thesis are LR, DTR, RFR, GBR, XGBoost, MLPRegressor, LightGBM and NARX. The first seven models were used as a benchmark for the time series model, NARX. The performance metrics used to evaluate the models'

performances are RMSLE, RMSE, MAPE, and MAE.

Based on the results and the statistical tests, we can conclude that for the Online Market Dataset among the evaluated regression models, the RFR shows the highest R^2 value (0,63), indicating the best performance among the models listed. The DTR also performs well, followed closely by XGBoost and LightGBM. LR shows the lowest R^2 value, indicating that it might not be the best choice for this dataset.

Moreover, by examining the results for The Food Demand Forecasting dataset from Kaggle, we can conclude that among the evaluated regression models, the LightGBM model exhibits the best performance, having the lowest RMSE (138,45) and highest R^2 (0,87), as well as the lowest MSE (19.169,66) and a low MAE (75,51). The RFR also performs exceptionally well, with an RMSE of 150,38, R^2 of 0,86, MSE of 22.614,96, and MAE of 68,94. Other models like XGBoost and MLPRegressor also show strong performance but are slightly outperformed by LightGBM and RFR. LR demonstrates significantly lower performance compared to the other models, with higher RMSE, MSE, and MAE, and a lower R^2 value.

The NARX models were evaluated on the test dataset, yielding an RMSE of 280,7, an R^2 of 0.58, an MSE of 78.792,58, and an MAE of 117,23. Compared to other models, the NARX model shows moderate performance. It performs significantly better than the LR model, which has a higher RMSE of 346,4, a lower R^2 of 0,25, a higher MSE of 119.990,55, and a higher MAE of 195.82. However, the NARX model is outperformed by models such as the DTR, RFR, GBR, XGBoost, MLPRegressor, and LightGBM. For instance, the LightGBM model, which has the best performance among the evaluated models, exhibits an RMSE of 138,45, an R^2 of 0,87, an MSE of 19.169,66, and an MAE of 75,51, all significantly better than the corresponding metrics for the NARX model. Similarly, the RFR, another top performer, has an RMSE of 150,38, an R^2 of 0,86, an MSE of 22.614,96, and an MAE of 68,94. In summary, while the NARX model is an improvement over LR, capturing the underlying patterns in the data more effectively, it is less accurate and effective compared to the more advanced models like LightGBM and RFR.

In addition, we trained the NARX model one more time with different features. The features we used in the new training are "week", "center_id", "meal_id", "checkout_price", "base_price", "emailer_for_promotion", "homepage_featured". The performance of the model was better with the new features. In the test dataset, we obtained R^2 of 0,58 in the first training and with the new feature the R^2 has moved up to 0,65. We also managed to decrease the overall prediction error.

The improvement in the performance of the NARX model can be attributed to the selection of more relevant and comprehensive features in the second model. In the first model, the features used were "category," "cuisine," "center type," "checkout_price," and "week." These features, while useful, provided a limited perspective on the factors influencing the number of orders. In contrast, the second model included a more extensive set of features: "week," "center_id," "meal_id," "checkout_price," "base_price," "emailer_for_promotion," and "homepage_featured." The inclusion of "center_id" and "meal_id" introduced specific identifiers for the distribution centers and meals, allowing the model to capture unique patterns and variations associated with different centers and meal types. The "base_price" feature offered additional pricing context, which, along with "checkout_price," enabled the model to better understand the pricing dynamics and their impact on orders. Furthermore, the marketing features "emailer_for_promotion" and "homepage_featured" provided insight into the influence of promotional activities on customer behavior. In the feature selection step, we have already seen that these features have moderate correlation with the number of sales, so it was necessary to use these for the NARX model. These features likely contributed to capturing the temporal and promotional effects more effectively. By incorporating a broader range of features that directly impact the target variable, the second model could more accurately capture the underlying patterns and trends in the data, leading to better predictive performance as evidenced by the improved RMSE, R^2, MSE, and MAE metrics. The comprehensive feature set in the second model allowed for a more nuanced understanding of the factors driving the number of orders, thereby enhancing the model's accuracy and robustness.

In addition, the second NARX model demonstrated notable improvements over the initial model, reflecting the impact of incorporating a broader and more relevant set of features. The second model achieved an RMSE of 257,48, an R^2 of 0,65, an MSE of 66.298,52,

and an MAE of 110,64. When compared to other machine learning models, the second NARX model shows a marked enhancement in performance. It significantly outperforms the Linear Regression (LR) model, which has an RMSE of 346,4, an R^2 of 0,25, an MSE of 119.990,55, and an MAE of 195,82. Despite these improvements, the second NARX model still lags behind the more sophisticated models such as DTR, RFR, GBR, XGBoost, MLPRegressor, and LightGBM. The LightGBM model exhibits superior performance with an RMSE of 138,45, an R^2 of 0,87, an MSE of 19.169,66, and an MAE of 75,51, significantly surpassing the metrics of the second NARX model. Similarly, the RFR also outperforms the NARX model, with an RMSE of 150,38, an R^2 of 0,86, an MSE of 22.614,96, and an MAE of 68,94. In summary, while the second NARX model shows considerable improvements and is more effective than the LR model, capturing underlying patterns more accurately, it still does not match the predictive accuracy and effectiveness of the more advanced models like LightGBM and RFR.

The study highlighted significant variability in model performance across the two datasets, underscoring the importance of understanding the nature of the data and the specific characteristics of the forecasting problem. The Online Market Dataset, characterized by a diverse set of features including historical sales data and product attributes, posed challenges in capturing non-linear relationships and complex interactions. In this context, the RFR achieved the highest R^2 value, demonstrating its robustness and ability to handle large datasets with numerous variables. Other advanced models like LightGBM and XGBoost also performed well, leveraging their boosting techniques to manage the dataset's high dimensionality.

In contrast, the Kaggle Food Demand Forecasting Dataset, which included features such as center_id, meal_id, pricing information, and promotional activities, presented a different set of challenges. LightGBM emerged as the top performer for this dataset, with an RMSE of 138,45 and an R^2 of 0,87. Its capability to handle large-scale data efficiently and capture complex interactions made it particularly effective. The Random Forest Regressor (RFR) also demonstrated strong performance, highlighting its versatility across different types of datasets. The improvement in the NARX model's performance with the inclusion of additional features such as center_id and meal_id further underscored the critical role of feature engineering in enhancing model accuracy.

The consistent performance of advanced models like RFR and LightGBM across both datasets indicates their robustness and adaptability to various data characteristics. However, simpler models such as LR showed limited effectiveness in capturing the complex patterns inherent in both datasets, reaffirming the need for more sophisticated algorithms in demand forecasting. This variability in performance across datasets emphasizes the necessity for a comprehensive evaluation process in practical applications. It involves careful selection, tuning, and benchmarking of models based on the specific characteristics of the dataset and the forecasting problem at hand.

One aspect that this study focused on was perishable and non-perishable goods. In the realm of food demand forecasting, the differentiation between perishable and non-perish able goods are critical due to the inherent characteristics and varying shelf lives of these product types. Perishable goods, which include categories such as fruits, vegetables, and organic products, have a limited shelf life and are more susceptible to spoilage. This unique nature necessitates a more precise and specialized approach to demand prediction to ensure freshness and minimize waste.

The results from our study clearly indicate that separating perishable goods and modeling them independently significantly enhances predictive accuracy. For instance, after isolating the perishable goods and creating a smaller, focused dataset, the performance metrics of several models improved markedly. The RFR, for example, saw its R^2 value rise from 0,63 to 0,79. Similarly, XGBoost's R^2 value increased from 0,57 to 0,75. These improvements underscore the effectiveness of targeted modeling strategies for perishable goods.

Perishable goods often exhibit distinct demand patterns influenced by factors such as seasonality, weather conditions, and promotional activities. By focusing solely on these goods, models can more effectively capture these unique trends without the noise introduced by non-perishable items. Combining perishable and non-perishable goods in a single model introduces high variability due to their differing shelf lives and demand cycles. Isolating perishable goods reduces this variability, allowing models to make more accurate predictions.

In summary, this thesis has contributed to the understanding of demand forecasting by evaluating a range of models and highlighting the importance of feature selection and algorithm choice. The insights gained from comparing the two datasets reveal that while advanced models like LightGBM and RFR show great promise, the effectiveness of these models can vary based on the nature of the data. Ongoing research and innovation are essential to further refine these techniques and ensure their effective application in diverse real-world contexts. By incorporating a broader range of relevant features and leveraging advanced algorithms, practitioners can achieve more accurate and reliable demand predictions, thereby enhancing decision-making processes in various industries.

REFERENCES

Kantasa-Ard, M. Nouiri, A. Bekrar, A. Ait el Cadi, and Y. Sallez, “Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand,” International Journal of Production Research, vol. 59, no. 24, pp. 7491 7515, 2021.

Analytics Vidhya. (2020). Food Demand Forecasting [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/1138962

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification And Regression Trees. In Routledge eBooks. https://doi.org/10.1201/9781315139470

Chen, S., & Billings, S. A. (1989). Representation of nonlinear systems: the NARMAX model. International Journal of Control, 49(3), 1013-1032.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Dis - covery and Data Mining (pp. 785-794).

ERA-NET FOSC and SUSFOOD2 Common Call Opened | TÜBİTAK | Türkiye Bilimsel ve Teknolojik Araştırma Kurumu. (2024, June 14). TÜBİTAK | https://tubitak.gov.tr/en/node/11749.

European Commission. (2020). A Farm to Fork Strategy for a fair, healthy and environmentally friendly food system. Retrieved from https://food.ec.europa.eu/horizontal-%20%20topics/farm-fork-strategy_en

European Commission. (2020, March 20). A new Circular Economy Action Plan for a cleaner and more competitive Europe. Retrieved from https://ec.europa.eu/news room/growth/items/673419/en

EU actions against food waste. (n.d.). Food Safety. https://food.ec.europa.eu/safety/food-waste/eu-actions-against-food-waste_en

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting

machine. Annals of Statistics, 29(5), 1189-1232.

G. Dellino, T. Laudadio, R. Mari, N. Mastronardi, and C. Meloni, “A reliable decision support system for fresh food supply chain management,” International Journal of Production Research, vol. 56, no. 4, pp. 1458–1485, 2018.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Hinton, G. E. (1990). Connectionist learning procedures. Artificial Intelligence, 40(1-3), 185-234.

Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. Expert Systems With Applications, 76, 140–151. https://doi.org/10.1016/j.eswa.2017.01.022

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk, and A. Kozina, “Food demand prediction using the nonlinear autoregressive exogenous neural network,” IEEE Access, vol. 9, pp. 146 123–146 136, 2021.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18-22.

L. Nassar, I. E. Okwuchi, M. Saad, F. Karray, and K. Ponnambalam, “Deep learning-based approach for fresh produce market price prediction,” in 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020, pp. 1–7.

M. Aci and D. Yergök, “Demand forecasting for food production using machine learning algorithms: A case study of university refectory,” Tehnič̆kivjesnik, vol. 30, no. 6, pp. 1683–1691, 2023.

PedregosaFabian, VaroquauxGaël, GramfortAlexandre, MichelVincent, ThirionBertrand, GriselOlivier, BlondelMathieu, PrettenhoferPeter, WeissRon, DubourgVincent, VanderplasJake, PassosAlexandre, CournapeauDavid, BrucherMatthieu, Perrot-Matthieu, & DuchesnayÉdouard. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. https://doi.org/10.5555/1953048.2078195

Quinlan, J. R. (1992). Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (pp. 343-348). World Scientific.

SDG Indicators. (n.d.). https://unstats.un.org/sdgs

S. K. Panda and S. N. Mohanty, “Time series forecasting and modelling of food demand

supply chain based on regressors analysis," IEEE Access, 2023.

S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan, and K. Litsiou, "Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail," International Journal of Production Research, vol. 58, no. 16, pp. 4964–4979, 2020.

United Nations Environment Programme. (n.d.). Halving food waste and raising climate ambition: SDG 12.3 and the Paris Agreement. Retrieved from https://www.unep.org/news-and-stories/story/halving-food-waste-and-raising-climate-ambition-sdg-123-and-paris-agreement

Zhang, L., & Zhang, B. (2006, July). Hierarchical machine learning–a learning methodology inspired by human intelligence. In International Conference on Rough Sets and Knowledge Technology (pp. 28-30). Berlin, Heidelberg: Springer Berlin Heidelberg.