© August 2025 | IJIRT | Volume 12 Issue 3 | ISSN: 2349-6002

Food Demand Forecasting Analysis

Rajeshwaran K[1], Valentina R[2], Vignesh SR[3], Trishanthan R[4]
[1]*Asst, Professor, Dept. of ECE, Sri Ramakrishna Engineering College,Coimbatore, Tamilnadu, India*
[2,3,4] *Student, Dept. of ECE, Sri Ramakrishna Engineering College, Coimbatore, TamilNadu, India*

Abstract—**This project aims to forecast food demand for various outlets or regions using historical sales data and external factors like seasonality, promotions, and weather. Accurate forecasting helps reduce food waste and optimize supply chains. The dataset includes sales records, external variables, and time-series information. Exploratory Data Analysis (EDA) identifies seasonal trends, the impact of promotions, and regional demand patterns. Multiple machines learning models, including Linear Regression, Random Forest, Gradient Boosting, and ARIMA, SARIMA are used for predictions. Model performance is evaluated using metrics like MAE, RMSE, and R^2 Score. The best-performing model's forecasts are visualized using Tableau to support decision-making. The results indicate that Random Forest Regression is the best-performing model, demonstrating strong predictive accuracy with the lowest errors. Decision Tree Regression also showed moderate performance, followed by Gradient Boosting Regression and k-Nearest Neighbors Regression, which exhibited higher errors and lower accuracy. Based on this evaluation, Random Forest Regression is recommended for reliable food demand forecasting.**

Index Terms—**Deep learning, Machine learning, Time series analysis and demand analysis**

I. INTRODUCTION

Analyzing food demand forecasts involves estimating the future demand for food products based on historical data, market dynamics, consumer behavior, economic influences, and climate conditions. This analysis is essential for businesses, supply chain professionals, and governments to facilitate effective food production, distribution, and consumption. Reliable demand forecasting enables companies to streamline supply chains, decrease inventory expenses, curtail waste, and guarantee product availability, which in turn enhances customer satisfaction and competitive edge. Furthermore, forecasting aids in strategic planning across various departments. The finance team assesses costs, profits, and investment requirements, while the marketing department utilizes forecasts to strategize campaigns and measure their effectiveness. The purchasing team formulates investment strategies, and operations are responsible for the timely acquisition of resources. By having dependable demand forecasts, organizations can improve their operational efficacy and profitability. [1]

The composition of time series Before delving into forecasting tasks in the following sections, it is important to describe time series. In demand time series, several patterns can be recognized. These patterns may not always be present,but when they do appear, they can be formalized and projected using appropriate methods to anticipate future sales. Trend: this refers to a long-term increase or decrease in the data. It can be linear or exhibit a more complex behavior. Seasonal: this entails periodic variations in demand driven by seasonal factors. The pattern maintains a fixed frequency. Cyclic: a cycle occurs when fluctuations in demand demonstrate a non-precise periodic behavior. This typically arises from economic cycles.

Operational dimensions of supply chains When addressing a demand forecasting challenge, it is crucial to consider the supply chain dynamics to tackle the issue using the most effective tools and methods. The forecasting environment within the supply chain can be characterized by three primary operational dimensions (Syntetos et al., 2016): length, depth, and time. Length: forecasting is required at various locations within the supply chain. Demand at the retail level generates demand at the subsequent upstream link (distributor), which will then respond by placing an order to the next link (manufacturer), and so forth. Length can be defined as the dimension encompassing all links in the supply chain. The nature of the demand will differ depending on the position in the supply chain. Depth: forecasting is employed for different levels of decision-making, ranging from inventory management to strategic planning. Depth is defined as the level of detail required for the information. This level of detail revolves around several key aspects: products, suppliers, customers, and locations. Time: this dimension encompasses both operational choices (like time buckets and forecast horizon) and data characteristics (such as data history, demand frequency, etc.).

Forecasting methods Within quantitative techniques, three disti nct forecasting methods can be recognized based on the data utilized for generating the forecast (Hyndman and Athanasopoulos, 2018): Time series methods: the forecast is generated using previous recorded demand as input. The majority of the commonly used classical techniques fall into this category. Causal (or explanatory) methods: this category utilizes predictor variables such as promotional activities, advertising efforts, and product characteristics to forecast upcoming demand. In other words, demand is linked to the input variables, which may or may not have causality ties. Predictor variables may either be dynamic (depending on time) or static. [2]

Food demand analysis involves several key steps. First, it begins with data collection, where historical data on food Data on consumption, sales, production, weather patterns, and economic indicators is collected. The subsequent phase involves identifying trends, where elements such as seasonal variations, changes in consumer preferences, and the influence of economic or environmental factors on demand are determined. Following this, modeling methods such as statistical analysis or machine learning are employed to forecast future demand based on the recognized trends and external influences. Next, scenario analysis is conducted to assess various future possibilities, including economic shifts or climate changes, and their potential effects on food demand. Ultimately, conclusions are drawn to guide businesses and governments, providing advice on optimizing inventory, mitigating supply chain risks, and planning production or imports to align with the anticipated demand. This thorough strategy assists organizations in proactively managing and preparing for future food requirements while maintaining the resilience and efficiency of the food supply chain. [3]

Machine learning (ML) is essential in demand forecasting within supply chain management (SCM), providing improved predictive accuracy that aids in making better decisions. By utilizing ML algorithms, organizations can make more strategic and informed decisions, which boosts revenue generation and enhances stock valuation. Research over the last ten years has significantly focused on sales demand forecasting in the food sector, with a systematic review demonstrating the advantages of ML techniques for predicting sales across various retail settings, including confectionery shops, grocery stores, and restaurants. ML models, particularly those employed for forecasting food demand, surpass traditional methods by minimizing human error and increasing prediction dependability. This progress assists companies in optimizing their inventory levels, decreasing stockouts, and improving overall supply chain effectiveness. Models like SARIMA (Seasonal AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory networks) are recognized as particularly efficient in retail SCM, delivering enhanced forecasting abilities compared to standard approaches.

These methodologies enable businesses to gain a better understanding of demand trends, ensuring they can fulfill consumer needs while keeping costs low. The passage addresses how machine learning (ML) models are utilized in demand forecasting for the food industry, specifically focusing on predicting fruit imports and fresh produce orders. One particular study employed a neural network-based approach to forecast the yearly import volume for fruits, while another explored several forecasting models— LSTM networks, Support Vector Regression (SVR), Random Forest Regression (RFR), Gradient Boosting Regression (GBR), Extreme Gradient Boosting Regression (XGBoost/XGBR), and ARIMA—to fine-tune daily order quantities for fresh produce and avert stockouts in a retail outlet on a college campus.

II. LITERATURE SURVEY

This survey explores recent studies that apply ML models for demand forecasting across various sectors, including retail, food production, and restaurant management. It evaluates the effectiveness of LSTM networks, optimized hyperparameter models, and hybrid approaches in enhancing forecast accuracy. By examining the challenges and advancements in ML-based forecasting, this survey provides insights into how businesses can leverage data-driven decision-making to optimize their supply chains and improve operational efficiency. Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. International Journal of Logistics Research and Applications, 25(2), 119-142.In his 2022 article, "Machine Learning Demand Forecasting and Supply Chain Performance," Javad Feizabadi addresses the challenges upstream firms face due to demand information distortion in multi-stage supply chains, leading to operational inefficiencies. He investigates the potential of advanced demand forecasting methods, specifically machine learning (ML) techniques, to mitigate these issues and enhance supply chain performance'Feizabadi develops hybrid demand forecasting models that integrate Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX) and Neural Networks. These models incorporate both time series data and explanatory factors to improve forecast accuracy. The study applies and evaluates these ML-based methods in the context of a steel manufacturer producing functional products.The findings reveal statistically significant improvements in supply chain performance when utilizing ML-based forecasting methods compared to traditional approaches. This underscores the effectiveness of machine learning techniques in enhancing forecast accuracy and operational efficiency within supply chains. The study also discusses the theoretical and practical implications of adopting ML-based forecasting methods, highlighting their potential to address demand information distortion and improve overall supply chain performance

Predictive Analytics for Demand Forecasting–A Comparison of SARIMA and LSTM in Retail SCM. Procedia Computer Science, 200, 993-1003.In their 2022 study, "Predictive Analytics for Demand Forecasting–A Comparison of SARIMA and LSTM in Retail SCM," Falatouri, Darbanian, Brandtner, and Udokwu examine the effectiveness of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) neural networks in forecasting demand within retail supply chain management (SCM). The authors highlight the critical role of accurate demand forecasting in enhancing supply chain efficiency and responsiveness. They note that traditional statistical methods like SARIMA have been widely used for time series forecasting but may struggle with capturing complex, non-linear patterns in data. In contrast, LSTM networks, a type of deep learning model, are designed to learn and remember long-term dependencies, potentially offering improved forecasting accuracy for intricate demand patterns.To assess the performance of these models, the researchers conduct a comparative analysis using real-world retail sales data. They evaluate each model's accuracy in predicting future demand and their ability to handle the inherent complexities of retail data, such as seasonality and trends.The study's findings reveal that the LSTM model outperforms the SARIMA model in terms of forecasting accuracy. This suggests that deep learning approaches like LSTM can more effectively capture the non-linear and complex relationships present in retail demand data, leading to more reliable forecasts. The authors conclude that incorporating advanced predictive analytics methods, particularly LSTM networks, can significantly enhance demand forecasting processes in retail SCM, ultimately contributing to better decision-making and operational efficiency.

Garre, A., Ruiz, M. C., & Hontoria, E. (2020). Application of MachineLearning to support production planning of a food industry in the context of waste generation under uncertainty. *Operations Research Perspectives*, 7, 100147.In their 2020 article, "Application of Machine Learning to Support Production Planning of a Food Industry in the Context of Waste Generation Under Uncertainty," Garre, Ruiz, and Hontoria explore how machine learning (ML) techniques can enhance production planning in the food industry, particularly by addressing waste generation amid uncertain conditions. The authors recognize that production planning in the food sector is complicated by factors such as perishable raw materials, fluctuating demand, and variable production yields, all of which contribute to uncertainty and increased waste. To tackle these challenges, they propose integrating ML models capable of analyzing historical production data to predict and manage uncertainties more effectively. The study demonstrates that incorporating ML into production planning allows for more accurate forecasting of variables like demand and production yields. This improved accuracy enables better decision-making, leading to optimized resource utilization and reduced waste. The authors conclude that ML applications can significantly enhance the efficiency and sustainability of food production processes by mitigating the impacts of uncertainty.

Zhang, H., & Lin, A. (2020, December). Research on Demand Analysis Model of Hot Product in Food Industry. *In 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)* (pp. 595-602). IEEE.In their December 2020 conference paper titled "Research on Demand Analysis Model of Hot Product in Food Industry," Zhang and Lin present a demand analysis model tailored for trending products within the food industry. The authors recognize that accurately forecasting demand for popular food items is crucial for optimizing inventory management, reducing waste, and enhancing customer satisfaction.To address this, they propose a model that integrates data mining techniques with machine learning algorithms to analyze factors influencing the demand for these trending products. By leveraging historical sales data, market trends, and consumer behavior patterns, the model aims to provide more precise demand forecasts.The study involves applying this model to real-world data from the food industry, demonstrating its effectiveness in capturing demand fluctuations and improving forecasting accuracy. The authors conclude that their approach offers valuable insights for food industry stakeholders, enabling better decision- making in production planning and inventory control.

Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022). Machine Learning Based Restaurant Sales Forecasting. *Machine Learning and Knowledge Extraction*, 4(1), 105- 130.In their 2022 article, *"Machine Learning Based Restaurant Sales Forecasting,"* Schmidt, Kabir, and Hoque explore the application of machine learning (ML) models to predict restaurant sales, aiming to optimize employee scheduling and inventory management. Using three years of sales data from a mid-sized restaurant, they applied feature engineering and selection techniques to identify relevant variables. The study compared various ML models, including Long Short- Term Memory (LSTM) and Gated Recurrent Unit (GRU), with traditional forecasting methods. Results showed that RNN models, particularly LSTM and GRU, outperformed traditional models in capturing complex patterns and improving prediction accuracy. Additionally, incorporating trend and seasonality adjustments enhanced forecast reliability. The authors concluded that advanced ML techniques offer significant potential for improving operational efficiency in the restaurant industry.

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP*, 88, 580-583.In their 2020 study, *"Restaurants Store Management Based on Demand Forecasting,"* Tanizaki, Hoshino, Shimmura, and Takenaka propose a demand forecasting model to optimize restaurant inventory management by aligning inventory orders with actual customer demand. Using historical sales data and machine learning techniques, the model accounts for factors like time of day, day of the week, and seasonal variations to accurately predict demand. The implementation of this model in a restaurant chain resulted in reduced overstocking and understocking, minimizing waste and improving customer satisfaction. The authors conclude that integrating demand forecasting into restaurant management practices enhances operational efficiency and resource utilization.

Abbasimehr, H., Shabani, M., & Yousefi, M.(2020).An optimized model using LSTM network for demand forecasting. *Computers & industrial engineering*, 143, 106435.In their 2020 article, *"An Optimized Model Using LSTM Network for Demand Forecasting,"* Abbasimehr, Shabani, and Yousefi propose a demand forecasting method based on multi-layer Long Short-Term Memory (LSTM) networks. The authors employ a grid search method to automatically select the best forecasting model by evaluating different combinations of LSTM hyper parameters for a given time series. This approach aims to enhance the accuracy of demand forecasts by leveraging the inherent capabilities of LSTM networks to capture complex temporal patterns in data.

III. METHODOLOGY

The goal of food demand forecasting is to predict future needs for food items, allowing businesses, governments, and various stakeholders to make informed decisions that enhance food production, supply chains, and delivery. By effectively forecasting demand, organizations can better manage inventory levels, minimize waste, and prevent stock shortages. This analysis also aids in the strategic allocation of resources, reducing production and transport costs, while ensuring customer satisfaction by providing the right products when needed. Furthermore, demand forecasting informs policy-making related to food security and sustainable agricultural practices, helping to control price volatility and lessen environmental impact. In summary, food demand prediction promotes a more efficient, cost- effective, and resilient food system that can adapt to evolving consumer demands and market dynamics.

IV.PROPOSED METHOD

This initiative seeks to predict food demand for particular outlets or regions by examining historical sales figures alongside external influences such as seasonality, promotions, and weather conditions. Precise forecasting aids in minimizing food waste and enhancing supply chain efficiency within the food sector. The dataset comprises sales information, external variables, and time-series data to discern trends and patterns.

The project initiates with Exploratory Data Analysis (EDA) to reveal crucial insights, including seasonal patterns, the effects of promotions, and variations in regional demand. A variety of machine learning algorithms are employed for forecasting, such as Linear Regression, Random Forest, Gradient Boosting, and ARIMA time-series models. The Effectiveness of these models is assessed using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score to identify the most precise model.

The model that performs the best is utilized to produce demand forecasts, which are subsequently illustrated in Tableau to facilitate decision-making and enhance inventory and supply chain management. This strategy improves operational efficiency, decreases waste, and results in more accurate food demand forecasting.

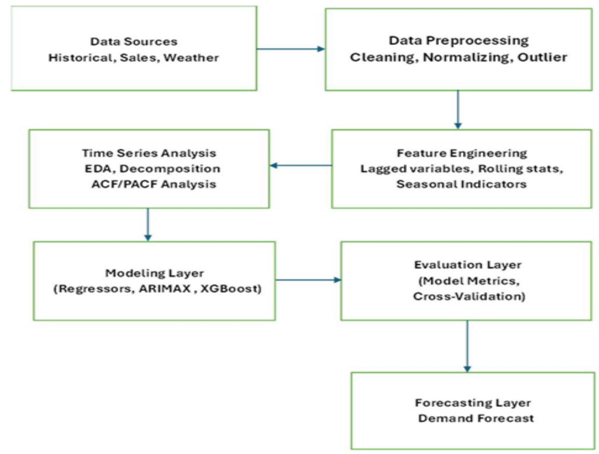


Figure 1

A.Data Collection

Historical sales information, which encompasses weekly demand, seasonal trends, promotional activities, weather influences, and additional significant factors like pricing and customer profiles, is gathered.

The dataset comprises five files: fulfilment_center_info.csv, meal_info.csv, sample_submission.csv, train.csv, and test.csv. The fulfilment_center_info.csv file includes details about fulfillment centers, such as their locations, types, and operational areas. The meal_info.csv outlines information regarding meals, detailing their categories and cuisines. The primary historical dataset, train.csv, contains weekly food demand data along with features such as checkout price, base price, promotional indicators, and the number of orders. The test.csv file is similar to the training dataset but does not include the num_orders column, making it ideal for forecasting future demand. Finally, the sample_submission.csv provides a format for submitting predictions, featuring placeholders for the anticipated number of orders. Collectively, these datasets facilitate the creation of predictive models aimed at accurately forecasting food demand.

The demand dataset contains 456,548 records and 9 columns, representing historical food demand data with features like prices, promotions, and the number of orders. The fulfillment center information dataset has 77 records and 5 columns, providing details about different centers, including their location, type, and operational area. The meal information dataset consists of 51 records and 3 columns, offering information on meal categories and cuisines. The test dataset has 32,573 records with 8 columns, similar to the training data but without the actual number of orders, used for making predictions. Additionally, a separate test dataset with 32,573 records and 2 columns might represent a simplified or filtered version, possibly for submission purposes.

B. Training Data

The demand dataset contains 456,548 entries and 12 attributes, detailing food demand across various fulfillment centers and meal types. The id column uniquely identifies each entry, ranging from 1,000,000 to 1,499,999. The dataset spans weeks 1 to 145, with an average week number of approximately 74.77. The columns center_id and city_code exhibit a broad distribution, ranging from 10 to 186. The average operational area (op_area) measures 56.6 square kilometers, with a minimum of 23 and a maximum of 93 square kilometers. The meal_id values fall between 0.9 and 7, indicating a variety of meal categories and cuisines. Regarding pricing, the checkout_price fluctuates between 1062 and 2956, with an average of 2024.34. Likewise, the base_price ranges from 55.35 to 866.27, averaging 354.16, suggesting significant price variability due to discounts and promotions. Promotions were utilized infrequently, as reflected by the low average values of 0.081 for emailer_for_promotion and 0.109 for homepage_featured. The total number of orders (num_orders) displays considerable variation, ranging from 13 to 24,299, with an average of 261.87 and a standard deviation of 395.92. This variation underscores the influence of factors such as promotions, pricing, and regional preferences on food demand.

C. Test Data

The dataset contains 32,573 records with 12 columns, representing information related to food demand, fulfillment centers, and meal details. The id column, which uniquely identifies each record, has values ranging from approximately 1,000,085 to 1,499,996. The data covers weeks between 146 to 155, with an average week number of around 150.48.

The center_id and city_code values exhibit considerable variation, with averages of 81.9 and 601.5, respectively. The operational area (op_area) ranges from 23 to 93 square kilometers, with a mean of 56.7 sq. km. Meal-related information shows the average meal_id is around 2032.07, with prices ranging widely. The checkout_price varies from 67.9 to 1113.62 and the base_price from 89.24 to 1112.62, reflecting the impact of promotions and discounts. Promotions through emails (emailer_for_promotion) and homepage features (homepage_featured) are relatively rare, with averages of 0.066 and 0.081, indicating minimal promotional activity. Notably, the num_orders column remains constant at 123456 across all records, suggesting a placeholder or erroneous data for this field. Further data cleaning and verification are recommended before analysis or model training.

D. Data Visualization

In the data visualization depicting the relationship between the week number and the number of orders, the trend of weekly food demand over the period of 145 weeks is illustrated. The graph typically shows fluctuations in demand, with the number of orders ranging between 0 and 25,000. Notably, visual inspection reveals that orders exceeding 5,000 are considered outliers, suggesting occasional spikes in demand. These outliers could be attributed to specific events such as promotions, seasonal changes, or festivals that lead to a temporary surge in food orders. Identifying such anomalies is essential for understanding consumer behavior and refining the prediction model. By using appropriate data preprocessing techniques like outlier removal or scaling, the accuracy of the model can be enhanced. Additionally, recognizing patterns of demand variation over time helps businesses plan inventory, manage supply chains effectively, and prevent food wastage

E. Interpretation of the Plot

The visual representation (Figure 2) will display the variations in the number of orders throughout the weeks.

Observations such as recurring trends or spikes during particular weeks may become apparent.

Outliers with significantly high order volumes will likely be visible, suggesting potential promotional activities or unexpected surges in demand.

This visual examination helps to comprehend demand patterns, facilitating enhancements to the prediction model by addressing irregularities or incorporating relevant external factors.

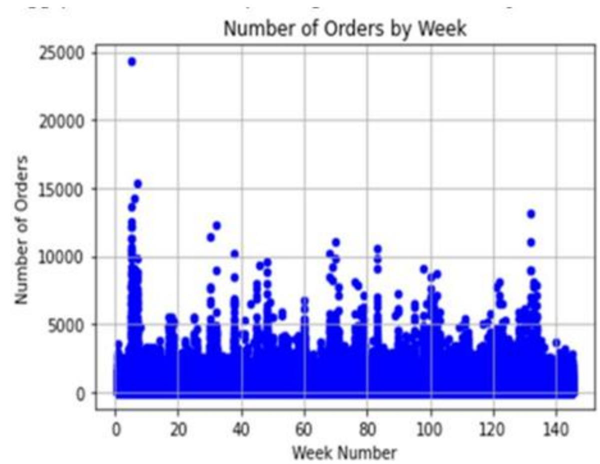


Figure:2

A category countplot (Figure 3) is a bar graph that illustrates how many times each category appears in a specific dataset.

In this case, the countplot depicts the distribution of various food categories within the dataset, revealing a total of 14 unique categories.

The visualization clearly indicates that beverages are the most favored category, reflecting a greater consumer preference or more frequent orders for drinks compared to other food items. This insight can be valuable for businesses as it helps them understand which categories contribute most to overall sales. High demand for beverages might suggest opportunities for further promotions, bundling, or inventory adjustments to meet the anticipated demand. Conversely, categories with lower order counts might indicate a need for targeted marketing or menu adjustments.

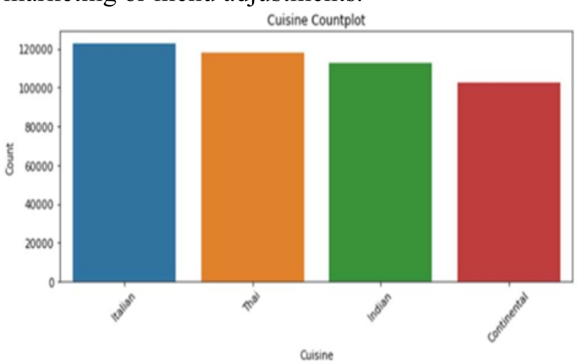


Figure:3

Additionally, identifying category trends over time can assist in forecasting future demands and optimizing supply chain management. The countplot serves as a simple yet effective visualization to derive actionable insights for business strategy and inventory planning.

Count of Cuisine Regarding cuisine, there are four types offered, with Italian being the most popular.

F. Interpretation of the Plot:

- The countplot(Figure 4) will reveal which cuisines are the most popular based on the number of orders.
- It provides insights into customer preferences, helping businesses tailor their food offerings.
- For example, if certain cuisines are significantly more popular, promotions or targeted marketing can be applied to those items.
- Conversely, cuisines with lower demand may require analysis to identify possible reasons, such as poor visibility or pricing issues.

Overall, this visualization aids in making data-driven decisions for product management and optimizing the food supply chain.

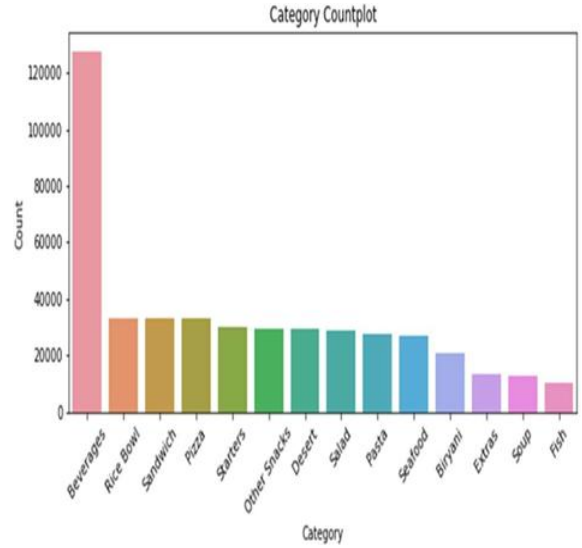


Figure :4

G. Heat map Analysis:

Key Feature Correlations with Number of Orders The heatmap analysis shows that the op area, cuisine, checkout price, base price, emailer for promotion, and homepage featured are the features most strongly correlated with the target variable, the number of orders

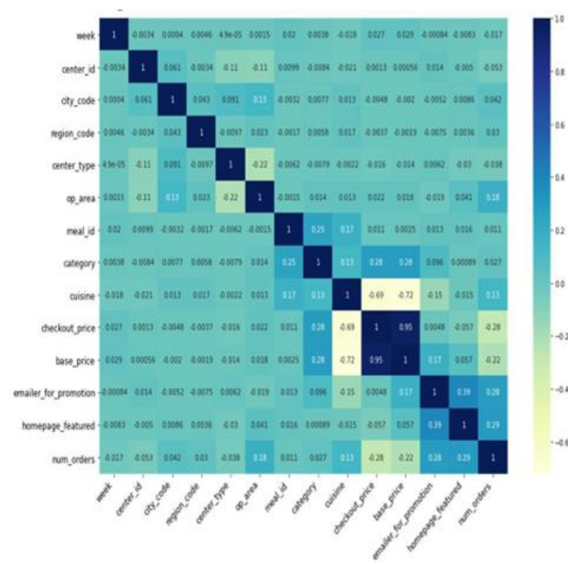


Figure:5

V.MODEL BUILDING PROCESS

The model building process for food demand prediction followed a systematic and iterative approach over five stages, aiming to enhance model accuracy and performance.

First Iteration: The model was developed using the default settings with no data preprocessing or feature adjustments. This baseline model provided an initial understanding of the data's predictive capacity and served as a reference for evaluating improvements in subsequent iterations.

Second Iteration: Outliers were addressed by removing orders above 5000, which were identified as extreme values based on data visualization. Removing outliers helped to reduce noise and prevent the model from being biased by abnormal demand spikes, resulting in improved prediction accuracy.

Third Iteration: One-hot encoding was applied to categorical variables like "center_id", "city_code", "center_type", "meal_id", "category", and "cuisine". This transformation converted categorical data into a numerical format, making it suitable for machine learning algorithms while preserving the information carried by categorical relationships.

Fourth Iteration: A combination of the previous iterations was implemented by applying outlier removal and one-hot encoding together. This holistic approach refined the data further, contributing to a more stable and accurate model.

Fifth Iteration: The final iteration focused on feature selection by using a heatmap correlation analysis to identify the most relevant features strongly correlated with the target variable (number of orders). Only these significant features were retained, reducing model complexity, avoiding overfitting, and enhancing computational efficiency. Overall, this structured iterative approach led to a progressively refined model, ensuring better generalization on test data. The use of feature engineering techniques like outlier removal, one-hot encoding, and correlation analysis played a key role in improving the prediction accuracy of the food demand forecasting model.

Model	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration5
k-NN	0.5228	0.5606	0.1111	0.1332	0.2599
DT	0.5672	0.6072	0.6165	0.6285	0.2157
RF	0.7291	0.7565	0.7452	0.7603	0.4067
GB	0.5862	0.6138	0.6574	0.6819	0.4357

Table :1

First Iteration: Used default settings with no modifications. The code effectively splits the dataset into training and testing sets for building a food demand prediction model using a time-based approach. Data from weeks 1 to 117 is used as the training set, while data from weeks 118 to 145 serves as the test set. This ensures that the model learns from historical data and is evaluated on future data, maintaining the chronological integrity essential for time series forecasting. The features (X) are extracted by dropping the num_orders column, while the target variable (y) consists of the actual number of orders. This separation allows the model to predict the demand based on relevant input variables. By preventing data leakage and simulating a real- world scenario where future demand is predicted using past trends, this method ensures accurate and reliable model evaluation. Furthermore, using a large test set provides comprehensive insights into the model's performance, making it a robust approach for food demand forecasting.

The performance of the food demand prediction model using different regression algorithms is evaluated using R-Squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Among the four models, the Random Forest Regression demonstrated the best performance with an R-Squared Score of 0.7291, indicating that it explains approximately 72.91% of the variance in food demand. It also recorded the lowest MAE of 88.39 and an MSE of 36,288.65, suggesting that its predictions are more accurate and have lower error rates compared to the other models.

The k-Nearest Neighbors (KNN) Regression achieved an R- Squared Score of 0.5228, with an MAE of 114.59 and an MSE of 63,926.28. While it provides moderate predictive accuracy, its higher error rates indicate it may struggle with complex relationships in the data. The Decision Tree Regression performed slightly better with an R^2 of 0.5671, a lower MAE of 113.81, and an MSE of 57,987.33. While it captures some patterns in the data, it may be prone to overfitting, reducing its generalization ability.

The Gradient Boosting Regression showed an R-Squared Score of 0.5862 and a higher MAE of 124.81 compared to the other models, with an MSE of 55,438.75. Although gradient boosting is generally effective for reducing bias, its relatively lower performance here could be due to model hyperparameters, insufficient tuning, or sensitivity to noise. Overall, the Random Forest Regression emerges as the most reliable and accurate choice for food demand prediction, making it the preferred model for further refinement and deployment

Second Iteration: Outliers were removed by excluding orders above 5000.

The performance of the food demand prediction model using four different regression algorithms is presented through key evaluation metrics: R-Squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). Among these models, the Random Forest Regression outperformed the others with an R^2 score of 0.7565, meaning it explains approximately 75.65% of the variance in food demand. It also achieved the lowest MAE of 86.69 and the smallest MSE of 27,999.98, indicating its superior accuracy and generalization ability.

The Decision Tree Regression followed with an R^2 score of 0.6072, reflecting moderate accuracy. It had a relatively low MAE of 110.87 and an MSE of 45,163.08, demonstrating its capability to capture patterns effectively, though it may be prone to overfitting. The k-Nearest Neighbors (KNN) Regression had an R^2 score of 0.5604, with an MAE of 112.18 and an MSE of 50,548.56, indicating moderate performance. While it showed improvement over the previous iteration, it remains less effective compared to ensemble methods.

The Gradient Boosting Regression achieved an R^2 score of 0.6138 and an MSE of 44,409.89, with a higher MAE of 120.42. Despite its general efficiency in handling complex data, its higher error rate suggests the need for further hyperparameter tuning.

Overall, Random Forest Regression remains the most reliable model for predicting food demand due to its lower prediction errors and higher explanatory power. This makes it the preferred choice for deployment in a real-world food supply chain management system.

Third Iteration: One-hot encoding was applied to the "center_id", "city_code", "center_type", "meal_id", "category", and "cuisine" columns.

The evaluation results for the food demand prediction model using four different regression algorithms show clear distinctions in performance. Among these models, Random Forest Regression stands out as the best performer, achieving an R-Squared (R^2) Score of 0.7452, indicating it explains approximately 74.52% of the variation in food demand. It also recorded the lowest Mean Absolute Error (MAE) of 86.87 and a Mean Squared Error (MSE) of 34,138.78, demonstrating high prediction accuracy and reliable performance. Decision Tree Regression followed with an R^2 Score of 0.6165, an MAE of 108.91, and an MSE of 51,380.24. While it captures patterns effectively, its tendency to overfit results in lower generalization performance compared to Random Forest. Gradient Boosting Regression performed moderately, with an R^2 Score of 0.6574, an MAE of 115.92, and an MSE of 45,892.87. Despite its robust learning capabilities, the higher MAE suggests further tuning may be necessary for better accuracy.

On the other hand, k-Nearest Neighbors (KNN) Regression performed poorly, with an R^2 Score of 0.1111, indicating it explained only around 11.11% of the variance in food demand. It also had the highest MAE of 172.69 and an MSE of 119,088.48, suggesting significant prediction errors. This poor performance may be attributed to its sensitivity to noise and inability to capture complex patterns in the data.

Fourth Iteration:

Combined the procedures from the second and third iterations.The evaluation results for the food demand prediction model using different regression algorithms highlight the strengths and weaknesses of each approach. Random Forest Regression remains the best performer with an R-Squared (R^2) Score of 0.7603, meaning it explains around 76.03% of the variance in food demand. Additionally, it has the lowest Mean Absolute Error (MAE) of 85.84 and a Mean Squared Error (MSE) of 27,556.30, indicating strong predictive accuracy and reduced error margins.

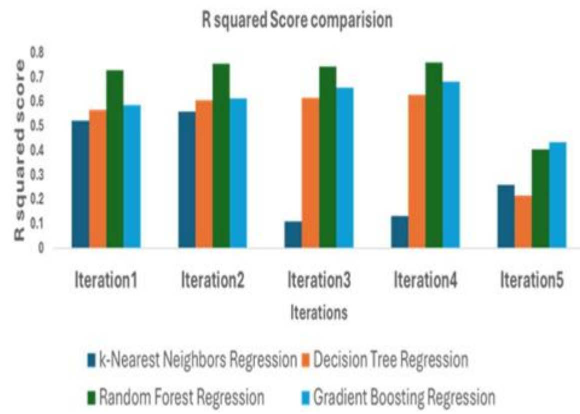
The Decision Tree Regression follows with a moderate performance, achieving an R^2 Score of 0.6285. While it has a relatively lower MAE of 107.61 and an MSE of 42,717.26 compared to other models, it may suffer from overfitting, which reduces its generalizability. Gradient Boosting Regression also performs reasonably well with an R^2 Score of 0.6819, an MAE of 112.85, and an MSE of 36,572.27. Although it captures patterns effectively, the higher MAE indicates that further tuning may be required for optimal accuracy. In contrast, the k-Nearest Neighbors (KNN) Regression performed poorly with an R^2 Score of 0.1333, signifying it explained only 13.33% of the variance. With the highest MAE of 169.77 and an MSE of 99,656.93, the KNN model struggled to generalize patterns in the dataset. Its poor performance may be attributed to its sensitivity to noise and the absence of clear neighborhood relationships in the data. Overall, Random Forest Regression is the most suitable choice for food demand prediction, offering reliable accuracy and stability. Further adjustments, like hyperparameter tuning or feature engineering, could improve the performance of other models, particularly Gradient Boosting and Decision Tree Regressions.

Fifth Iteration: Included only the features most strongly correlated with the target variable, number of orders, as identified in the heatmap analysis.

The food demand prediction model was evaluated using four different regression algorithms: k-Nearest Neighbors (KNN) Regression, Decision Tree Regression, Random Forest Regression, and Gradient Boosting Regression. Among these models, Random Forest Regression consistently demonstrated the best performance, achieving the highest R- Squared (R^2) Score of 0.7603, indicating it explained 76.03% of the variance in food demand. It also recorded the lowest Mean Absolute Error (MAE) of 85.84 and a Mean Squared Error (MSE) of 27,556.30, suggesting strong predictive accuracy.

Decision Tree Regression showed moderate accuracy with an R^2 Score of 0.6285, an MAE of 107.61, and an MSE of 42,717.26, but its tendency to overfit limited its generalization. Gradient Boosting Regression performed reasonably well, with an R^2 Score of 0.6819, an MAE of 112.85, and an MSE of 36,572.27. Despite its robust learning capabilities, further tuning could improve its accuracy.

On the other hand, KNN Regression exhibited poor performance with an R^2 Score of 0.1333, an MAE of 169.77, and an MSE of 99,656.93, indicating it struggled to capture the data's patterns. Overall, Random Forest Regression emerged as the most suitable model for food demand prediction due to its superior accuracy and lower error rates, making it the preferred choice for deployment.



R Squared error

Mean Squared Error

VI. CONCLUSION

In conclusion, food demand prediction plays a crucial role in optimizing supply chain management, minimizing food waste, and ensuring effective resource allocation in the food industry. By leveraging advanced time series forecasting models like ARIMA, SARIMA, and machine learning algorithms, businesses can accurately estimate future demand based on historical sales data and external factors. This enables better inventory management, reduced operational costs, and improved customer satisfaction. Furthermore, integrating predictive insights into decision- making processes helps companies respond proactively to market fluctuations, promotional impacts, and seasonal trends. Ultimately, food demand prediction serves as a strategic tool for achieving operational efficiency, reducing waste, and promoting sustainable practices across the food supply chain.

REFERENCES

[1] S. Kumar Panda, S. N. Mohanty: "Time Series Forecastingand Modeling of Food Demand Supply Chain"IEEE access,VOLUME 11, 2023.

[2] Vijay Kotu and Bala Deshpande "Data science concept and practice"second edition 2019.

[3] Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. International Journal of Logistics Research and Applications, 25(2), 119-142.

[4] Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Pre- dictive Analytics for Demand Forecasting–A Comparison of SARIMA and LSTM in Retail SCM.Procedia Computer Science, 200, 993-1003.

[5] Garre, A., Ruiz, M. C., & Hontoria, E. (2020). Application of MachineLearning to support production planning of a food industry in the context of waste generation under uncertainty. Operations Research Perspectives, 7, 100147.

[6] Zhang, H., & Lin, A. (2020, December). Research on Demand Analysis Model of Hot Product in Food Industry. In 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C) (pp. 595-602). IEEE.

[7] Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022). Machine Learning Based Restaurant Sales Forecasting. Machine Learning and Knowledge Extraction, 4(1), 105- 130.

[8] Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. Procedia CIRP, 88, 580-583.

[9] Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. Computers & industrial engineering, 143, 106435.