An Efficient Approach for Food Demand Forecasting Using an Ensemble Technique and Statistical Analysis

Dudla Anil kumar(✉), Bathula Thirupathi Rao, Bathini Rangaswamy, and Kagitha Meghana

Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram 521230, India
anil.dudla@gmail.com

Abstract. By utilizing machine learning, the current research offers a fresh approach to food demand forecasting in restaurants. It focuses on store-specific models that take into account a variety of variables, including location, weather, and events. Conventional methods frequently ignore store-specific subtleties in favor of relying exclusively on point-of-sale (POS) data. We provide a food demand forecasting model that utilizes machine learning methods, specifically XGBoost and CatBoost, to incorporate various data sources. The study goes into how the model was created, how real shop data was used to validate it, and how it was used to assess variables that affect consumer happiness, especially in the context of the fast-food trend. The study explores the variables influencing preferences for fast-food outlets and offers insights into consumer satisfaction and how it affects the fast-food restaurant industry as a whole. The benefits of the presented hypothesis, along with the algorithms and libraries used (Pandas, NumPy, Scikit-learn, and CatBoost) are described in depth. The study also emphasizes the significance of data visualization, ensemble techniques, hyperparameter adjustment, and other gradient-boosting libraries like XGBoost and CatBoost. Libraries for time series analysis are also thought to be useful for capturing data's temporal patterns. This all-encompassing method helps to create a strong and effective framework for food demand forecasting, which is crucial for restaurant management and optimization.

Keywords: Food Demand Forecasting · Machine Learning · XGBoost · CatBoost · Customer Satisfaction · Ensemble Technique · Hyperparameter Tuning · Data Visualization · Gradient Boosting · Restaurant Management and Optimization

1 Introduction

This study focuses on the critical area of restaurant business food demand forecasting [1]. While acknowledging the widespread application of machine learning in these kinds of forecasts, the research points out a critical flaw in current approaches: they pay insufficient attention to store-specific variables like location, weather, and events. Although

P. Pareek et al. (Eds.): IC4S 2024, LNICST 597, pp. 61–77, 2025.
https://doi.org/10.1007/978-3-031-77075-3_6

machine learning algorithms have shown to be efficient in managing extensive datasets and intricate patterns in consumer behavior, their implementation has frequently been broad, resulting in inadequate precision for forecasting demand for specific retailers. The food market [2] is projected to increase at an annual pace of 6.53% to produce revenue of US$10.07 trillion in 2024. With a market volume of US$1.77 trillion, confectionery and snacks is the largest sector. With a projected $1,630 billion in revenue in 2024, China is the largest market in the world [3]. In 2024, per capita revenue is $1,299.00 US. By 2024, it is anticipated that 4.3% of total income will come from online sales. In 2025, the market is expected to expand by 3.9%, with an average volume of 352.30 kg per person.

Previous investigations in the area have emphasized the significance of machine learning algorithms such as XGBoost and CatBoost [4], demonstrating their adaptability to handle a wide range of datasets. However, further research is needed to fully understand how to apply them in the context of food demand forecasting, particularly about store-specific details. To fill these gaps, this study offers a thorough method that includes specific variables in addition to machine learning techniques for forecasting [5]. The relationship between consumer happiness and food demand in the context of the fast-food trend is one important topic covered in this study. Previous research has demonstrated the complex relationship between consumer satisfaction and variables such as overall eating experience, service speed, and food quality. Understanding how important these components are, the study attempts to disentangle the complexities of fast-food trends [6] by presenting insights into the variables affecting consumer satisfaction and, in turn, impacting the demand for fast-food options.

1.1 Objectives

Using cutting-edge machine learning techniques and solving significant inadequacies in the present state of food demand forecasting are the primary goals of the project. First and foremost, the main objective is to create a food demand forecasting model [7] that is accurate and efficient while going beyond conventional techniques by taking into account store-specific variables. Current methods frequently over-rely on Point of Sale (POS) data [8], ignoring specific attributes like location, meteorology, and neighborhood activities. The project intends to improve the accuracy and usefulness of food demand forecasts by including these variables in the machine learning framework [9]. This would give restaurant management insightful information to help them make better decisions.

The subsequent objective of the study is to investigate and assess the variables affecting consumer satisfaction in the context of the fast-food industry. Fast-food businesses rely heavily on customer happiness, yet little is known about the complex interactions between the several elements that go into making a satisfied customer. Finding and examining these variables will help to clarify issues with meal quality, service time, and general eating experience. The study's comprehension of the nuances of customer satisfaction not only advances the conversation about consumer behavior in general but also offers fast-food restaurants practical advice on how to customize their menus and services, therefore affecting industry trends. By pursuing these two goals, the study hopes to improve our knowledge of food demand forecasts [10] from a theoretical and practical standpoint.

2 Literature Survey

The report's debut aims to discuss the crucial topic of restaurant food demand forecasting, which is a vital topic in the fast-paced and fiercely competitive hospitality sector [11]. This study's basis is the realization that conventional methods, which mostly depend on Point of Sale (POS) data, would not be sufficient to produce precise and specific forecasts. Previous research in this area has mostly concentrated on using machine learning to estimate food demand, while store-specific variables like location, weather, and events have received less attention. The research now in publication has documented several attempts to use machine learning methods to forecast food demand. Numerous studies have shown how effective algorithms [12] are at managing huge datasets and identifying intricate patterns present in consumer behavior. Still, a significant gap exists as these methods sometimes fail to take into account the distinct environment of specific establishments, leading to less-than-ideal forecasting accuracy. To close this gap, this study suggests a methodology that methodically adds store-specific variables into the forecasting process in addition to machine learning methods.

According to Harshini [13], food waste and deterioration pose serious issues for businesses that sell food. AI systems aim to decrease waste by forecasting sales and demand for raw materials. This paper proposes a demand forecasting system that uses stacking techniques to anticipate customer numbers, dish sales, and the required quantity of raw materials. Restaurants may cook meals while cutting down on waste if their MAE metric is between 0.4 and 0.7. Demand forecasting, according to Akyuz [14], is crucial for the retail industry's operational efficiency, cost reduction, and eradication of stock-out problems. It boosts revenue and sales, but it also results in unhappy and churning customers. A few things that influence prediction accuracy include social events, competitor behavior, trends, seasonality, and promotional effects. An innovative heuristic approach for the ensemble methodology was implemented at SOK Market, a 4000-store hard discount chain. Raju [15] and colleagues' study provides a strong basis for forecasting demand in the steel industry through the application of regression ensemble models, feature selection, cross-validation, data transformation, and preprocessing. The framework incorporates bagging, boosting, and stacking models in addition to reference models from machine learning approaches such as SVR, ELM, and MLP. The ensemble technique reduces decision-making risk and improves model performance, making it suitable for one-month demand forecasting. Hyperparameters are set using the grid search method.

Seyedan [16] offers a three-step, data-driven, cluster-based demand forecasting method for the retail sector that uses time-series analysis and Bayesian model averaging to segment consumers based on RFM criteria. Ribeiro [17] investigates how well regression ensembles may predict agricultural commodity prices. The comparisons of bagging, boosting, and stacking ensembles were conducted using monthly time series data from Parana, Brazil. Priyadarshi [18] the goal of this research is to use performance analysis to determine which forecasting model is best for a certain vegetable at the retail stage. Jayapal [19] and the team suggest employing machine learning algorithms to forecast precise food orders for Italian eateries to lower waste, cut expenses, and raise customer happiness through better staff management and a reduction in food and raw material loss.

The function that consumer satisfaction plays within the fast-food trend [20] is one of the major factors taken into account in this research. Previous research has demonstrated that an awareness of the elements of customer satisfaction is essential to the success of fast-food businesses [21]. By assessing these variables, we can identify the components that contribute to customers' overall pleasure as well as provide insights into their preferences. According to the research currently in publication, customer satisfaction [22] is a complex phenomenon that is impacted by a variety of aspects, from the entire eating experience to the quality of the food and speed of service. To fully understand the complexities of fast-food trends and consumer happiness, this research attempts to dig further into these aspects.

Hoxha [23] by using historical data from 1975 to 2019, this study suggests a machine learning stacking ensemble approach to forecast Turkey's transportation energy demand. With the Extra Tree Regressor and ADABoostRegressor acting as meta-regressors, the model employs hyperparameter adjustment and multicollinearity reduction. When all characteristics are taken into account, the model's R-squared value is 0.99; when only two features are taken into account, its accuracy is 0.98. This study can help establish more effective plans for controlling the energy consumption of transportation and encouraging environmentally friendly urban growth. Aci [24] presents a demand forecasting model for a university library using Machine Learning techniques, utilizing 18 models for optimal prediction performance. Kilimci [25] by using deep learning models, time series analysis, and machine learning, this work creates an intelligent demand forecasting system. It is the first to combine these techniques with an ensemble strategy enhancement. The algorithm exhibits a notable improvement in accuracy when tested using actual data from Turkey's SOK Market. Ganesan [26] says that demand forecasting in agriculture management is crucial for global economic growth. This paper presents a comparative analysis using machine learning algorithms to develop a Time series prediction model for primary crops like paddy and wheat.

Moreover, prior studies in the domain have demonstrated the adaptability of machine learning algorithms like XGBoost and CatBoost in managing a variety of datasets. These algorithms are already commonplace across many disciplines because of their shown effectiveness in tasks like regression and classification [27]. However, a sophisticated comprehension and adaptation are needed for their implementation in the context of food demand forecasting, especially with a focus on store-specific peculiarities. To address the unique problems presented by the complexities of restaurant operations, this research expands on the foundations set by previous analyses and suggests an effective ensemble approach that combines the benefits of XGBoost and CatBoost [28].

3 Data Gathering

To acquire comprehensive datasets that capture the many aspects impacting food demand and consumer satisfaction within the fast-food business, comprehensive datasets must be gathered for this research project (as shown in Fig. 1). The datasets' richness is derived from a variety of sources, including as event logs, store-specific data, meteorological data, and Point of Sale (POS) data. POS data, which forms the basis of the demand forecasting model [29], offers transactional details including item sales, time of purchase, and consumer preferences. Store-specific data includes specifics.

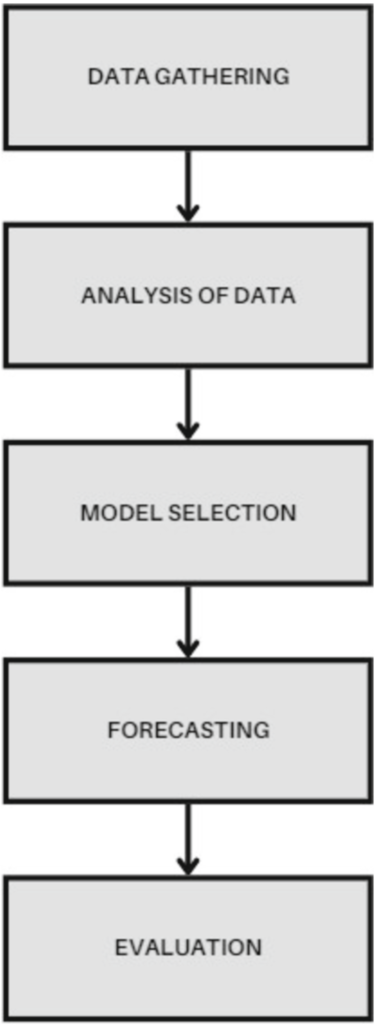

Fig. 1. Steps for collection of Data and Evaluation

about each outlet's location, and size which makes it possible to create customized forecasting models. To take into consideration the effects of the environment on food intake, weather data is included, including temperature, precipitation, and seasonal trends [30].

The timeline of the data collection is carefully considered, encompassing a suitably representative period, to ensure the integrity and usefulness of the datasets [31]. To address missing values, outliers, and inconsistencies, the datasets are cleaned and preprocessed, guaranteeing the precision and dependability of the ensuing studies. Utilizing a variety of data sources also supports the research's focus on developing a comprehensive and contextually rich forecasting model that can take into account the complex relationships between many elements influencing food demand and customer satisfaction in fast-food restaurants.

3.1 Train to Learn Information

The “Train to Learn Information” dataset [32] has 456,548 items in total, each with the following attributes: ‘id,’ ‘week,’ ‘center_id, “meal_id,’ ‘checkout_price,’ ‘base_price,’ and ‘num_orders.’ Each unique entry in the collection includes details about certain weeks, meal and center codes, pricing details (base and checkout rates), and the number of orders that correspond with each. The ‘id’ of each record uniquely identifies it, while the ‘week’ denotes the week of data entry. Part of the structured dataset that facilitates machine learning model training is ‘num_orders’, which is the goal variable to be predicted based on the given attributes. This dataset contains important data, such as meal and center identities and pricing information, which helps develop models that forecast food consumption [33] in a certain location.

3.2 Centre Information

There are 78 items in the “Centre Information” collection [34], and among its variables are “center_id,” “city_code,” “region_code,” “center_type,” and “op_area.” Every entry in the collection denotes a unique center and offers crucial information for examination. Each center is uniquely identified by its “Center_id,” and its geographic location is indicated by its “city_code” and “region_code.” By classifying centers according to their type, the ‘center_type’ offers information on the type of establishment. Furthermore, each center’s operational region is represented by ‘op_area’. In the larger framework of the dataset ecosystem [35], this dataset provides essential information for activities like center-specific demand forecasting and operational optimization. It also acts as a core element for understanding the distribution and characteristics of different centers.

3.3 Meal Information

There are 50 records in the “Meal Information” collection [36], and each one is defined by characteristics like “meal_id,” “category,” and “cuisine.“ Meal-specific information may be more easily associated across datasets thanks to the function of “meal_id,” which acts as a unique identifier for every meal. Meals are divided into a variety of categories by the ‘category’ property, including drinks, extras, soup, other snacks, salad, rice, appetizers, sandwiches, and more. This categorization offers a thorough grasp of the many kinds of offerings that are offered. Meals are categorized by their culinary origin using the ‘cuisine’ feature. This includes cuisines like Thai, Indian, Italian, Continental, and more. This dataset is essential for placing the unique qualities of each meal in context and enabling in- [37] with other datasets, especially when considering meal-related properties in the context of a larger dataset ecosystem.

4 Working Methodology

The research's functional approach is a methodical approach that includes developing a model, preparing data, and conducting a thorough analysis (as shown in Fig. 2). First, a thorough preprocessing is applied to the various datasets, such as "Train to Learn Information," "Centre Information," and "Meal Information." Maintaining the cleanliness and coherence of the datasets includes addressing missing values, and outliers [38], and standardizing formats. The incorporation of meteorological information, event logs, and store-specific data enhances the datasets and makes a more in-depth study possible. The creation of an effective model for predicting food demand is at the center of the process. XGBoost and CatBoost, two cutting-edge machine learning methods [39], are used to train the model using the enriched datasets. These algorithms help to create a strong ensemble approach because of their scalability and effectiveness in managing structured data. Tools like GridSearchCV and RandomizedSearchCV from sci-kit-learn or Optuna help with hyperparameter tuning, which fine-tunes the model's parameters for best results. Stacking, bagging, and boosting [40] are examples of ensemble algorithms that combine the advantages of separate models to increase forecasting accuracy.

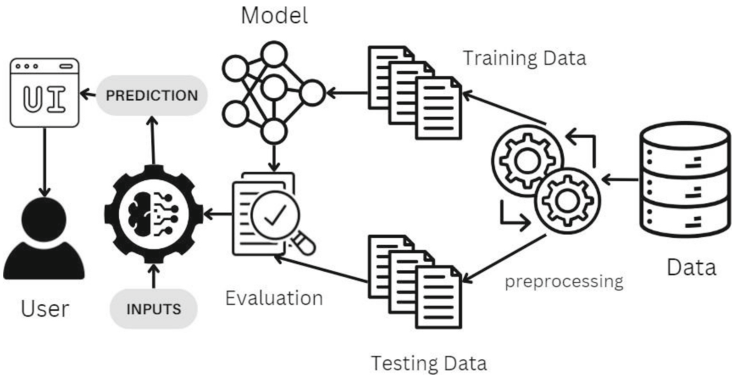

Fig. 2. Working Methodology of the system.

Meanwhile, the research investigates whether aspects of the fast-food trend affect consumer happiness. The outcomes of the model are examined and compared with measures of customer happiness, providing light on the connection between precise demand forecasting and satisfied customers. Tools for data visualization, such as Matplotlib, Seaborn, or Plotly [41], are essential for analyzing results and successfully conveying conclusions. The study also makes use of time series analysis libraries, such as stats models, prophets, or ARIMA [42], to find temporal trends that might affect the demand for food.

4.1 Data Processing

To ensure that the raw datasets are converted into a clear, organized format appropriate for training machine learning models, data preparation is an essential stage in this study.

Examining the datasets carefully and looking for anomalies, missing numbers, and contradictions is the first step. Missing values can seriously affect the performance of the model, thus handling them is essential. Depending on how much data is missing [43], several approaches, such as imputation or removal, may be used. The model may be used to fill in missing values for categorical variables like "center_type" and "cuisine," while mean or median imputation may be used for numerical features like order numbers or price. Another crucial component of data preparation is outlier detection. Values that are abnormally high or low might distort the model's comprehension and provide predictions that are off. To find and deal with outliers effectively, strategies including interquartile range (IQR) approaches and Z-score analysis [44] are applied. Depending on how extreme values affect the distribution of the dataset as a whole, they can either be eliminated or altered.

Normalization and standardization [45] are crucial processes that guarantee uniformity across various attributes with various scales. The 'checkout_price' and 'base_price' variables, for example, may have varying ranges, which might impact the efficiency of specific algorithms. Normalization scales the data to a certain range, usually between 0 and 1, whereas standardization scales the data to have a mean of 0 and a standard deviation of 1. Feature engineering is essential to increasing the richness of the dataset. In this study, new features that capture the subtleties of each entry are created through the integration of meteorological data, event logs, and store-specific information. For example, the 'checkout_price' and 'base_price' variables may be used to create a 'discount_percentage' feature, which would provide more details on pricing dynamics. In a similar vein, combining past order data to produce variables like "average_orders_per_week" advances our knowledge of demand trends.

Encoding categorical variables [46] is the last stage of data preparation. Machine learning models usually demand numerical inputs, and transformations must be performed on categorical variables such as 'center_type' or 'cuisine'. This is accomplished by using methods like label encoding, in which categories are given numerical labels, and one-hot encoding, in which binary columns are generated for each category. The type of variables and how they affect the model determine which encoding technique is best.

4.2 Predictive Analysis

In this research, predictive analysis entails using sophisticated statistical methods to project future food demand based on available historical data and pertinent characteristics. The primary aim is to create a resilient model that can precisely forecast the quantity of orders, an essential measure in the fast-food sector. A range of statistical techniques are utilized, with particular emphasis on machine learning algorithms like XGBoost and CatBoost [47], which are well-known for their efficiency in managing structured data and generating precise forecasts. The gradient boosting method used in this study, the XGBoost algorithm, is optimized for both speed and performance. To create a strong predictive model, it builds an ensemble of weak models using decision trees as base learners. Regularization strategies [48], managing missing values, and speed are some of the elements that contribute to the algorithm's efficiency; these attributes make it a good fit for big datasets with intricate interactions. The decision trees that form the basis of XGBoost's statistical framework are trained iteratively, with each tree learning from the mistakes of the one before it to produce an ensemble model that is both reliable and accurate.

The gradient boosting approach included in this prediction research, called CatBoost, was created expressly to deal with categorical data in a natural way. It is useful in situations where categorical variables are important since it doesn't need a lot of preprocessing [49]. Robustness against overfitting, high-performance execution, and efficient management of categorical feature interactions comprise its statistical basis. A crucial statistical component of predictive analysis is hyperparameter tuning, which is a methodical search for the ideal set of model parameters. In this study, tools such as GridSearchCV and RandomizedSearchCV [50] from sci-kit-learn or Optuna are used to optimize the XGBoost and CatBoost models' parameters. By guaranteeing that the models can adapt to a variety of datasets and increasing overall forecast accuracy, the statistical optimization method seeks to improve the models' generalization capabilities.

Ensemble techniques are an essential component of predictive analysis that combines several models' predictions for optimal performance. Boosting, stacking, and bagging are common group strategies. Using the diversity of multiple models to capture different facets of the underlying patterns in the data is the statistical justification for ensemble approaches [51]. Combining these many viewpoints strengthens the ensemble model and reduces overfitting, improving its prediction power. Libraries like Matplotlib, Seaborn, or Plotly provide data visualization [52], which aids in the statistical understanding of the outcomes of the predictive analysis. Time series plots, feature significance plots, and prediction vs. reality plots are a few examples of visualization approaches that help in understanding the performance of the model and spotting patterns and trends in the data. The statistical validation and interpretation of the results of the predictive analysis depend on this visual investigation.

4.3 XGBoost

Extreme Gradient Boosting, or XGBoost, is a potent and popular machine learning method that works well with structured data because of its scalability, efficiency, and efficacy. The XGBoost procedure consists of many phases, each of which helps to produce a strong ensemble model. XGBoost is a member of the gradient boosting algorithm class, which trains a series of weak learners (usually decision trees) [53] sequentially to fix the mistakes made by the earlier models. Starting with a weak model, more models are added one after the other to minimize the residual errors of the ensemble as a whole. The model's capacity to recognize intricate correlations and patterns in the data is improved by this recurrent training.

Decision trees are used by XGBoost as base learners [54]. The method chooses the optimal split at each node to minimize the loss function, building each tree in turn. The learning rate, regularization parameters, and tree depth are important hyperparameters that affect the model's complexity and capacity for generalization. Regularization techniques [55] are included in XGBoost to manage model complexity and avoid overfitting. Large coefficients are penalized and their influence on the model is limited by the addition of L1 and L2 regularization factors to the objective function. This improves the model's capacity to generalize effectively to new data and helps prevent the model from fitting noise in the training set.

XGBoost has features to deal with the dataset's missing values. Based on the data at hand, the algorithm decides how to handle missing values during training. This is especially helpful when working with datasets from the real world where missing values are typical. Missing value cases may be successfully included in the model-building process using XGBoost [56]. XGBoost uses distributed and parallel computing to achieve speed and efficiency in its architecture. The approach is very scalable and appropriate for huge datasets since it can make use of parallelism during tree formation. This efficiency is particularly helpful in situations where there is a shortage of computing power.

4.4 CatBoost

The gradient boosting method known as CatBoost, or Categorical Boosting [57], was created expressly to manage categorical features. For machine learning applications, its resilience against overfitting, fast handling of missing data, and distinct approach to categorical encoding make it a popular candidate. CatBoost stands out for its ability to handle categorical variables intuitively, without requiring a lot of preprocessing. CatBoost is capable of directly processing category characteristics, translating them into numerical representations during training, in contrast to other gradient boosting techniques. This makes the process of preparing the data easier, and it also makes it possible for the algorithm to precisely extract the information that is inherent in categorical variables [58].

A key problem with machine learning models is overfitting, which CatBoost addresses with built-in techniques. It uses a mix of feature permutations during training to determine feature significance and depth regularization, which restricts the depth of the individual trees in the ensemble. By using these strategies, the model is better able to generalize to new, untested data and is kept from fitting noise in the training set. CatBoost is designed to operate with maximum efficiency and performance. To increase computational performance during training, it combines oblivious trees, symmetric tree structure, and ordered boosting. Large datasets are also supported by the technique, which makes use of parallelization [59] to speed up model training. CatBoost has the same built-in ability to handle missing values in the dataset as XGBoost. The algorithm takes into account the best course of action for handling missing data in light of the information at hand during training. This is especially useful for real-world datasets where missing values are frequent since CatBoost can efficiently include these occurrences in the model-building process.

4.5 User Interface

The creation of interactive web apps is made easier by the simplified method of creating a user interface using Streamlit. Installing Streamlit [60] using a package manager is the initial step, which enables users to rapidly configure their development environment. The Python script then starts importing the required libraries, which include scikitlearn for machine learning features and pandas for data processing, among other things. Loading and preparing the data is the second step after setting up the development environment.

The *st. File_uploader* function in Streamlit makes it easier for users to integrate datasets into their applications by enabling direct data uploading. After that, Pandas may be used to do various data manipulation activities, such as cleaning and changing data to meet the needs of the application.

Building the user interface components is the third stage, which entails the process's central activity. Furthermore, without requiring complicated HTML or CSS, Streamlit's interactive widgets [61] such as sliders, buttons, and choose boxes allow the building of dynamic and user-friendly interfaces. Integrating data visualizations and insights straight into the interface is the main goal of the fourth stage. With Streamlit, users may generate interactive graphs and charts by utilizing well-known plotting libraries such as Matplotlib or Plotly [62]. The application may easily incorporate these visuals, improving user experience and providing a clear comprehension of the data or machine learning outcomes.

5 Results

The outcomes derived from the all-encompassing analytics framework [63] comprise both descriptive and predictive insights, offering a thorough comprehension of the data and possible future patterns. A thorough overview of past data is provided by descriptive analytics (as shown in Fig. 3), which also highlights important characteristics, patterns, and trends (as shown in Fig. 4) in the dataset. Graphical representations like histograms, scatter plots, and time series charts [64] are used to show the distribution of variables, correlations, and temporal patterns (as shown in Fig. 5) using data visualization tools like Matplotlib, Seaborn, or Plotly. Stakeholders may more easily understand complicated data patterns thanks to these visuals, which improve interpretability.

Predictive analytics (as shown in Fig. 6), which goes beyond descriptive analytics, uses cutting-edge machine learning models like XGBoost and CatBoost to greatly improve the outcomes. To anticipate future trends and especially, food demand in the context of this research—these models are trained on previous data. The predictive models are adjusted using ensemble techniques and hyperparameter tweaking to maximize their accuracy (as shown in Fig. 7) and generalization potential. In the fast-food business, predictive analytics helps with strategic decision-making, inventory management, and resource allocation by enabling stakeholders to predict future results. Apart from forecasting abilities, the outcomes also include statistical metrics assessing the model's effectiveness. The prediction accuracy of the model is measured using metrics like R-squared, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) [65]. These metrics provide a data-driven assessment of the predicted performance by giving stakeholders a quantitative assessment of how well the model matches the actual observed values [66]. Furthermore, displaying and visualizing the findings depends heavily on the user interface created using Streamlit. Streamlit facilitates the smooth incorporation of analytical results, both predictive and descriptive, into an interactive and intuitive interface. Stakeholders get real-time access to data trends, interactive visualizations, and insights. The use of a user-centric strategy improves the accessibility of findings, rendering them more actionable and enabling the fast-food business to make well-informed decisions.

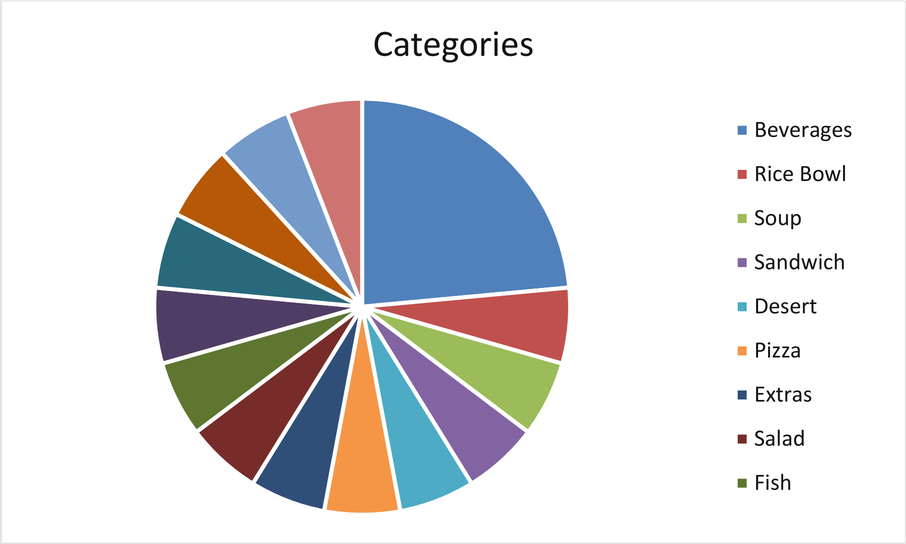

Fig. 3. Descriptive Analytics of the data.

Fig. 4. Visualization of the various attributes.

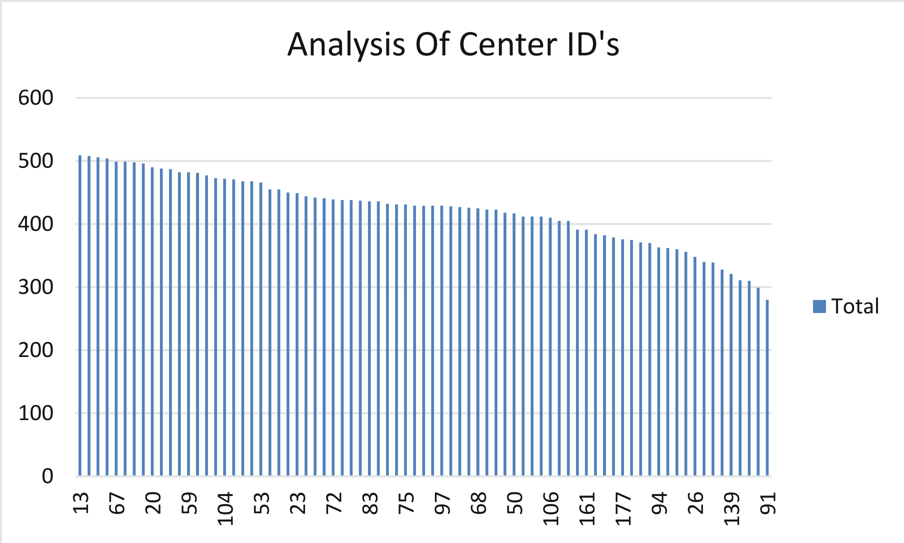

Fig. 5. Column graph of the data.

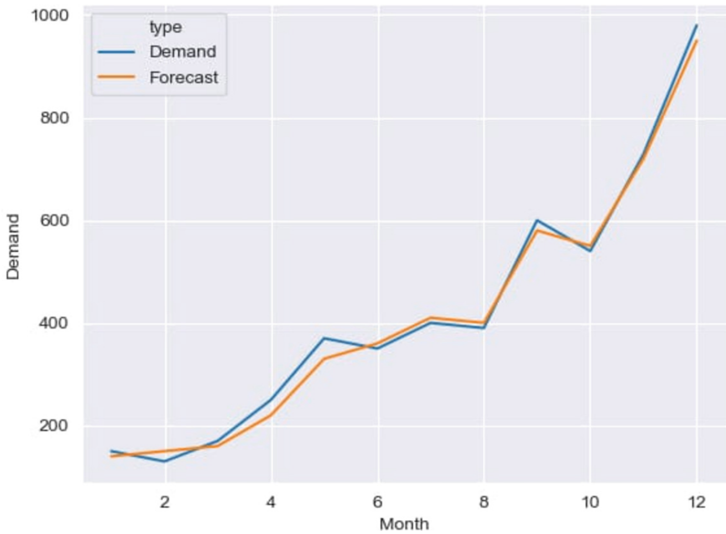

Fig. 6. Demand vs Monthly Forecasting

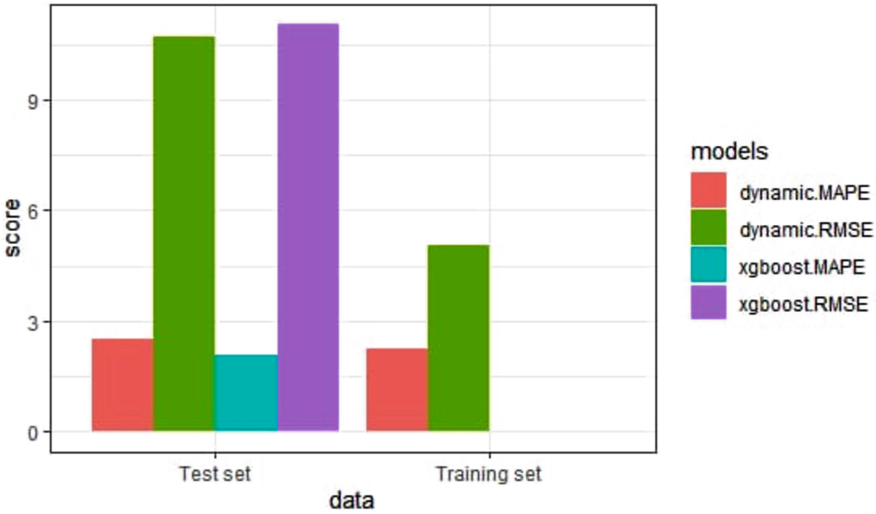

Fig. 7. Different Models vs Accuracy Score on Test and Training Data

6 Conclusion

The current research combines store-specific variables with machine learning techniques like XGBoost and CatBoost to provide an effective and customized method of food demand forecasting in the fast-food business. In addition to demonstrating precise predictive analytics, the findings offer insightful information about the complex relationships that exist between store-specific factors, consumer happiness, and general fast-food trends. The Streamlitdeveloped user interface makes these findings more accessible by providing stakeholders with an easy-to-use platform to examine and make use of descriptive and predictive analytics. The comprehensive approach utilized in this study advances our knowledge of food demand trends and promotes data-driven decision-making in the ever-changing fast-food sector.

References

1. Lasek, A., Cercone, N., Saunders, J.: Restaurant sales and customer demand forecasting: literature survey and categorization of methods. In: Smart City 360°: First EAI International Summit, Smart City 360°, Bratislava, Slovakia and Toronto, Canada, 13–16 Oct 2015. Revised Selected Papers 1 (2016): pp. 479–491 (2015)
2. Hancock, J., Khoshgoftaar, T.M.: Performance of catboost and xgboost in medicare fraud detection. In:19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE (2020)
3. Carbonneau, R., Laframboise, K., Vahidov, R.: Application of machine learning techniques for supply chain demand forecasting. Eur. J. Oper. Res. **184.3**, 1140–1154 (2008)
4. Zota, A.R., Phillips, C.A., Mitro, S.D.: Recent fast food consumption and bisphenol A and phthalates exposures among the US population. In: NHANES, 2003–2010. Environmental Health Perspectives, vol. 124.10, pp. 1521–1528 (2016)

5. Da Veiga, C.P., et al.: Demand forecasting in food retail: a comparison between the Holt-Winters and ARIMA models. WSEAS Trans. Bus. Econ. **11.1**, 608–614 (2014)
6. Simon, R.: The ABCs of point of sales (POS) data. J. Bus. Forecast. **27.4** (2008)
7. Menon, A., et al.: A machine learning framework for programming by example. In: 11-International Conference on Machine Learning. PMLR (2013)
8. Flies, E.J., et al.: Forecasting future global food demand: a systematic review and meta-analysis of model complexity. Environ. Int. **120**, 93–103 (2018)
9. Chen, P.: Analysis of hotel management data based on fuzzy logic mining. In: Abawajy, J., Xu, Z., Atiquzzaman, M., Zhang, X. (eds.) 2021 International Conference on Applications and Techniques in Cyber Intelligence. ATCI 2021. LNDECT, vol. 81. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79197-1_47
10. Hunter, D.R., Lange, K.: A tutorial on MM algorithms. Am. Statist. **58.1**, 30–37 (2004)
11. Harshini, K., et al.: Enhanced demand forecasting system for food and raw materials using ensemble learning. In: 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV). IEEE (2021)
12. Akyuz, A.O., et al.: Ensemble approach for time series analysis in demand forecasting: ensemble learning. In: IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA) (2017)
13. Raju, S.M.T.U., et al.: An approach for demand forecasting in steel industries using ensemble learning. In: Complexity, pp. 1–19 (2022)
14. Seyedan, M., Mafakheri, F., Wang, C.: Cluster-based demand forecasting using Bayesian model averaging: an ensemble learning approach. Dec. Anal. J. **3**, 100033 (2022)
15. Ribeiro, M.H.D.M., dos Santos Coelho, L.: Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Appl. Soft Comput. **86**, 105837 (2020)
16. Priyadarshi, R., et al.: Demand forecasting at the retail stage for selected vegetables: a performance analysis. J. Model. Manage. **14.4**, 1042–1063 (2019)
17. Jayapal, S.: Food demand prediction using statistical and machine learning models. In: Diss. Dublin, National College of Ireland (2023)
18. Yahya, F., Zafar, R., Shafiq, S.: Trend of fast-food consumption and its effect on Pakistani society. Food Sci. Qual. Manage. **11**, 1–7 (2013)
19. Gilbert, G.R., et al.: Measuring customer satisfaction in the fast food industry: a cross-national approach. J. Serv. Market. **18.5**, 371–383 (2004)
20. Hom, W.: An overview of customer satisfaction models (2000)
21. Hoxha, J., et al.: Prediction of transportation energy demand in Türkiye using stacking ensemble models: methodology and comparative analysis. Appl. Energy **350**, 121765 (2003)
22. Aci, M., Yergök, D.: Demand forecasting for food production using machine learning algorithms: a case study of university refectory. Tehnički vjesnik **30.6**, 1683–1691 (2023)
23. Kilimci, Z.H., et al.: An improved demand forecasting model using deep learning approach and proposed decision integration strategy for the supply chain. Complexity (2019)
24. Ganesan, A., Paul, A., Jeon, G.: Demand forecasting for primary crop production using different machine learning algorithms: a comparative analysis. In: The 18th International Conference on Multimedia Information Technology and Applications (2022)
25. Torgo, L., Gama, J.: Regression by classification: advances in Artificial Intelligence: 13th Brazilian Symposium on Artificial Intelligence, SBIA 1996 Curitiba, Brazil, 23–25 Oct 1996 Proceedings 13. Springer Berlin Heidelberg, (1996)
26. Ahn, J.M., Kim, J., Kim, K.: Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. Toxins **15.10**, 608 (2023)
27. Rathipriya, R., et al.: Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. Neural Comput. Appl. **35.2**, 1945–1957 (2023)

28. Zhou, W., et al.: Predicting seasonal patterns of energy production: a grey seasonal trend least squares support vector machine. Expert Syst. Appl. **213**, 118874 (2023)
29. Bansal, H., Grover, A.: Leaving reality to the imagination: Robust classification via generated datasets. In: arXiv preprint arXiv:2302.02503 (2023)
30. Eastman, P., et al.: Spice, a dataset of drug-like molecules and peptides for training machine learning potentials. Sci. Data **10.1**, 11 (2023)
31. Foini, P., et al.: On the fore castability of food insecurity. Sci. Reports **13.1**, 2793 (2023)
32. Ali, S., et al.: A multi-center polyp detection and segmentation dataset for generalisability assessment. Sci. Data **10.1**, 75 (2023)
33. AlShammari, N.K., et al.: Post-stroke rehabilitation optimization & recommendation framework using telerobotic ecosystem: industry 4.0 readiness approach. J. Intell. Fuzzy Syst. Reprint **1–11**, 39 (2023)
34. Ammann, J., et al.: A review on policy instruments for sustainable food consumption. In: Sustainable Production and Consumption (2023)
35. Voizard, B., et al.: Preoperative evaluation of the depth of invasion in oral tongue squamous cell carcinoma: a systematic review and meta-analysis. Oral Oncol. **136**(2023), 106273 (2023)
36. Tong, H., Tortora, C.: Missing values and directional outlier detection in model-based clustering. J. Classific. 1–34 (2023)
37. Ming, W., et al.: A systematic review of machine learning methods applied to fuel cells in performance evaluation, durability prediction, and application monitoring. Int. J. Hydrogen Energy **48.13**, 5197–5228 (2023)
38. Mondal, S., Ghosh, S., Nag, A.: Brain stroke prediction model based on boosting and stacking ensemble approach. Int. J. Inform. Technol. 1–10 (2023)
39. Lavanya, A., et al.: Assessing the performance of python data visualization libraries: a review (2023)
40. Ospina, R., et al.: An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: methodology and case study in Brazil. Mathematics **11.14**, 3069 (2023)
41. Matsuba, M., et al.: Effectiveness of hierarchical Bayesian models for citizen science data with missing values: a case study on the factors influencing beach litter in Shimane Prefecture, Japan. Marine Pollut. Bull. **191**, 114948 (2023)
42. Yaro, A.S., Maly, F., Prazak, P.: Outlier detection in time-series receive signal strength observation using Z-score method with S n scale estimator for indoor localization. Appl. Sci. **13.6**, 3900 (2023)
43. Huang, L., et al.: Normalization techniques in training dens: Methodology, analysis, and application. IEEE Trans. Pattern Anal. Mach. Intell. (2023)
44. Nugroho, H., Utama, N.P., Surendro, K.: Smoothing target encoding and class center-based firefly algorithm for handling missing values in a categorical variable. J. Big Data **10.1**, 1–18 (2023)
45. Zhang, L., Jánošík, D.:Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approach. Expert Syst. Appl. **241**, 122686 (2024)
46. Kim, M., et al.: PartMix: regularization strategy to learn part discovery for visible-infrared person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
47. Werner de Vargas, Vi., et al.: Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. Knowl. Inform. Syst. **65.1**, 31–57 (2023)
48. Arauna, K.V.: comparison between stochastic gradient descent and support vector machine on e-commerce data using randomized search cv. In: Diss. Universitas Katholik Soegijapranata Semarang (2023)
49. Ibrahim, Z., Tulay, P., Abdullahi, J.: Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa. Environ. Sci. Pollut. Res. **30.2**, 3621–3643 (2023)

50. Ma, R., Sun, E.D., Zou, J.: A spectral method for assessing and combining multiple data visualizations. Nat. Commun. **14.1**, 780 (2023)
51. Costa, V.G., Pedreira, C.E.: Recent advances in decision trees: an updated survey. Artific. Intell. Rev. **56.5**, 4765–4800 (2023)
52. Asselman, A., Khaldi, M., Aammou, S.: Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact. Learn. Environ. **31.6**, 3360–3379 (2023)
53. Feng, D.-C., Ren, X.: Analytical examination of mesh-dependency issue for uniaxial RC elements and new fracture energy-based regularization technique. Int. J. Damage Mech. **32.3**, 321–339 (2023)
54. Hajek, P., Abedin, M.Z., Sivarajah, U.: Fraud detection in mobile payment systems using an XGBoost-based framework. Inform. Syst. Front. **25.5**, 1985–2003 (2023)
55. Qian, L., et al.: Employing categorical boosting (CatBoost) and meta-heuristic algorithms for predicting the urban gas consumption. Urban Climate **51**, 101647 (2023)
56. Bolland, M.J., Avenell, A., Grey, A.: Baseline categorical variables in spine randomized controlled trials (2023)
57. Wang, W., et al. :{TopoOpt}: co-optimizing network topology and parallelization strategy for distributed training jobs. In: 20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23) 2023
58. Boddapati, M.S.D., et al.: YouTube comment analysis using lexicon based techniques. In: International Conference on Cognitive Computing and Cyber-Physical Systems. Springer Nature Switzerland, Cham (2022)
59. Khorasani, M., Abdou, M., Fernández, J.H., Basics, S.: Web application development with streamlit: develop and deploy secure and scalable web applications to the cloud using a pure python framework, pp. 31–62. Apress, Berkeley, CA (2022)
60. Herda, G., McNabb, R.: Python for Smarter Cities: Comparison of Python libraries for static and interactive visualizations of large vector data. arXiv preprint arXiv:2202.13105 (2022)
61. Knoth, S., Schmid, W.: Control charts for time series: a review. Front. Statist. Qual. Control **7**, 210–236 (2004)
62. Chicco, D., Warrens, M.J., Jurman, G.:The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. PeerJ Comput. Sci. **7**, e623 (2021)
63. Boddapati, M.S.D., Desamsetti, S.A., Adina, K., Uppalapati, P.J., Murty, P.T.S.: Creating a protected virtual learning space: a comprehensive strategy for security and user experience in online education. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (eds.) Cognitive Computing and Cyber-Physical Systems, IC4S 2023 (2024)
64. Dudla, A.K., Rao Atthuluri, M., Shaik, P.S., Yalamanchili, V.S.B.: An efficient approach for analyzing reviews using an ensemble technique. In: 9th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1576–1581. Coimbatore, India (2023). https://doi.org/10.1109/ICACCS57279.2023.10112729.(2023)
65. Kumar, D.A., Ezhilarasan, M.: Shufflenetv2: an effective technique for recommendation system in e-learning by user preferences. In: Morusupalli, R., Dandibhotla, T.S., Atluri, V.V., Windridge, D., Lingras, P., Komati, V.R. (eds.) Multi-disciplinary Trends in Artificial Intelligence. MIWAI 2023. LNCS, vol. 14078. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-36402-0_16(2023)
66. Kumar, D.A., Jyothi, U., Mohan, K.J., Mounica, V., Nageswari, A.: Kernel-based SVM classifiers for multi-disease forecasting: a meta-analysis. In: 4th International Conference on Smart Electronics and Communication (ICOSEC), pp. 841–847. Trichy, India (2023). https://doi.org/10.1109/ICOSEC58147.2023.10275963.(2023)