



# **International Journal of Multidisciplinary and Scientific Emerging Research (IJMSERH)**

**Volume 13, Issue 3, July-September 2025**

**Impact Factor: 9.274**



# Predicting Food Orders using Machine Learning Techniques

Nithish D<sup>1</sup>, Hemanth Kumar<sup>2</sup>

Student, Department of MCA, JNN College of Engineering, Shivamogga, Karnataka, India<sup>1</sup>

Associate Professor, Department of MCA, JNN College of Engineering, Shivamogga, Karnataka, India<sup>2</sup>

**ABSTRACT:** Predicting food orders accurately is a perpetual challenge in the restaurant business. This work will explore providing accurate food order prediction through Machine Learning classifications. An accurate forecasting of food orders is key to achieving less food waste, a proactive inventory forecast and happier clients. Increased population and digitalization have significantly increased food demand. Restaurants, University canteen services, catering services around the world have one conundrum, predicting how much of a food order will come in for any given day. If they fail to show accuracy in predicting food orders, they risk running short of food or excess food waste and costs associated with wasted food and employee management and business profitability. The work utilized a dataset of 5000 entries obtained with variable parameters, such as item name, previous year record, weather, time slot, delivery mode, discount status, holiday status, day of the week, status of the week end and temperature in degree. This work concluded based on available features, LightGBM appears to a very strong model for predicting food orders. The successful implementation into web application as demonstrated a work able to real-world restaurants.

**KEYWORDS:** Machine Learning, Food Orders, LightGBM, Predicting orders.

## I. INTRODUCTION

In a fast digital world, especially in today's urban city's, people use meal delivery services which are Swiggy and Zomato, become a integrated into urban lifestyle quickly, with a million transactions occurring each day. It is vital to understand and predict consumer behavior to enhance user experience and better manage processes, which ultimately leads to better outcomes for businesses. This is where Machine Learning has emerged as a great asset, as it can help to make better decisions by leveraging the way it analyzes large datasets and discovers patterns. The informed decision-making that will now result, will lead to less waste of food and raw materials, more profit and provide fresh food to consumers, great-tasting food every time. The work Predicting Food orders Using Machine Learning Techniques, will utilize several advanced ML models which are- Random Forest, LightGBM, XGBoost, and Gradient Boosting to predict food orders of food items. The model will look at a dataset of 5000 values and their associated fields which will ideally take into account real life customer experiences based on past buying behaviours, which include items name, weather (rainny, sunny, spring), the previous years purchase record, time slot (Breakfast, Lunch, Dinner), delivery method (Din-in, Home), discount, holiday, day of week (Monday to Sunday) and temperature (In degree). The goal of this work will to create a successful model, to accurately identify order purchases of food items based on the presented conditions.

## II. LITERATURE REVIEW

R S Mallik et al. created a Machine Learning model to estimate retail goods sales online, with product reviews playing a key role in the forecast. RandomForest technique accuracy is good than DecisionTree and Support Vector Machine (SVM) [1]. S M Nazmuz Sakib work intension on sales in restaurants using two data sets of restaurants in Bangalore with features that include weather condition and public holidays, have impact on getting how much food orders and sales in restaurants [2]. Here he use algorithms: Linear Regression, Decision Tree, Support Vector MachineRegression and Light GBM. He evaluated using Mean Absolute Error, Light BGM performed good with the lowest error rate compare to other models. In [3], the authors try to provide restaurants owners with some guidance on staffing and food preparation with a better ML model to assist those restaurants owners. In this instance, the ML model captures numerous features, including date, day, weather, discount, and holidays. They discover that XGBoost performs good. Vignesh Bengalure Renukprasad's work improves predictive analysis for restaurants forecasting daily sales, based on training a model to predict total daily sales and individual sales. Here in this case the Multiple linear regression, K-

Nearest Neighbors, DecisionTree and RandomForest Algorithms was utilized for their work [4]. Batool Madani and Hussam Alshraideh in [5] use a dataset of the 388 consumers from Bangalore, which includes demographics like weather ,day and feedback. Here the C4.5 decision tree shows the highest accuracy at 91.67%. Among various models. They mainly focus on customer satisfaction and increasing repeat purchases. Hussam Mezher Merdas and Ayad Hameed Mousa in [6] aimed to minimize food waste and improve profit. They developed and compared multiple Machine Learning algorithms across two datasets, with Evaluation Metrics dataset 1 having a high correlation in that LASSO (Least Absolute Shrinkage and Selection Operator), SVR (Systemic Vascular Resistance), and Bagging Regressor have a high success rate, and dataset 2 having a low correlation in that Gradient Boosting, Random Forest Regressor, and Decision Tree have high accuracy. Nicolas Hubner et al. focuses on reducing food waste in bakeries by optimizing sales by forecasting with direct environmental impacts and indirect traditional ordering methods, using real-world case study data from 175 German bakeries [7]. They used a Recurrent neural network (RNN) for demand forecasting; it was selected for its ability to handle time series data to capture patterns in historical bakery sales, weather, and holiday effects. This model generates daily demand-adjusted sales forecasts to reduce overproduction, which leads to waste. Konstantin Posch et al. developed a model for forecasting daily sales of individual menu items in restaurants and staff canteens [8]. They propose a Bayesian modeling framework using generalized additive models to capture the key characteristics of real-world food sales time series, such as multiple strong seasonalities (daily, weekly, and yearly), trend shifts, and POS (point of Sale) data. Actually they have two models, which are Normal Bayesian GAM (Generalized additive model) and Negative Binomial Bayesian GAM. Takashi Tanizaki et al. designed a model for better store operations such as employee scheduling and ordering food and beverages by predicting customer demand through Machine Learning [9]. They used Random Forest Regression, to forecast the quantity of beverages which would be ordered and staff the size at restaurants by utilising internal data similar to Point of Sale and external data similar to weather patterns and event calendars. The authors in [10] presented ADEM (Autonomous Digital Experience Management), a web-based tool to predict how much food will be needed in institutional food courts (i.e., universities or hospitals). They have collected a 44-month dataset encompassing daily sales and menu data. They achieved 80.78% accuracy via 10-fold cross-validation. They developed using Microsoft SQL Server Analysis Service and ASP.Net (Active Server Pages Network Enabled Technologies) framework. K Lutoslawski et al. place their focus on utilizing artificial intelligence to predict food demand with the goal of decreasing food waste [11]. They take real sales data of the past 3 years of retail stores and apply a Nonlinear AutoregressiveNetwork with exogenous inputs and get accuracy up to 99.6%. The NARX model performed better than forecasting methods like ARIMA (Autoregressive Integrated Moving Average). Akshay Krishna et al. designed ML model with stores sales dataset, they observed that AdaBoost algorithm perfomed well with RMSE value 1350.72 compare to Gradient Boost with RMSE value of 1088.64, so AdaBoost is tunned [12]. Ayushi Vaidhy et al. aimed for ML-based analysis of Indian restaurants data from the Zomato platform and exploratory data analysis and ML algorithms to extract useful patterns and make predictions about restaurant sales and customer preferences [14]. ML models like Random Forest and Decision Tree have high accuracy. Nouran Nassibi et al. focus on demand forecasting in the food industry using Machine Learning models for food products like chocolate to improve supply chain efficiency [15]. They take data from a major Saudi Arabian chocolate company; by using Support Vector Machine and Long Short-Term Memory, it predicts which city has good performance.

### III. METHODOLOGY OF PROPOSED SURVEY

#### 3.1 Data Acquisition

The dataset used in this study consists of 5000 historical food order records derived from a simulated database reflecting online and offline food sales. Each record contains several independent parameters that influence client purchasing behavior. There are

- A. Item Name:** A particular item from a menu, or dish.
- B. Last Year Order:** Orders value over the same historical timeframe in the previous year.
- C. Weather Condition:** Categorical variable describing the weather the order is placed in including conditions such as autumn, rainy, spring, summer, sunny, winter.
- D. Time Slot:** Meal period classification, for example, morning, afternoon, evening.
- E. Delivery Mode:** Type of delivery e.g., dine in, take away, home delivery.
- F. Has Discount:** Boolean variable that describes the provision of discounts.
- G. Day of the Week:** Categorical variable for the days of week e.g., Monday to Saturday.
- H. Is Weekend:** Boolean variable flagging weekends.
- I. Temperature:** Numerical variable that gives the average temperature in the order timeframe.

### 3.2 Data Preprocessing

Data Preprocessing is performed to prepare the dataset for next phase model training. Missing and inconsistent data entries were processed by using methods of imputation with mean imputation for numerical attributes and mode imputation for categorical attributes. Categorical Encoding done by using standard scaler and one hot encoding performed feature scaling for numerical data and decoding categorical data with numerical format to allow for understanding machine learning models. The numerical features of previous year's order record and temperature were standardized through min and max scaling. At last the dataset was split into 80% training and 20% testing for models to be evaluated.

### 3.3 Model Selection and Training

Five supervised machine learning regression methods were chosen for experiments.

**A. XGBoost:** an open-source distributed machine learning software library that is based on gradient boosted decision trees, which is a supervised learning boosting algorithm using gradient descent.

**B. LIGHTGBM:** a free and open-source distributed gradient-boosting framework. It used decision tree algorithms and used in ranking, classification and all machine learning tasks.

**C. Gradient Boosting Regressor:** is a machine learning method that combines lots of weak predictive models usually decision trees, to obtain a predictive model with a higher accuracy.

**D. Random Forest Regressor:** a supervised learning algorithm that operates by constructing an ensemble of decision trees; it is used to predict numerical or continuous data.

**E. ElasticNet Regressor:** a regularized linear regression using the penalties of both Lasso (L1) and Ridge (L2) regression.

### 3.4 Model Evaluation

The trained models were evaluated using three statistical performance metrics, R2 (Coefficient of Determination) is a statistical term that measures the fraction of variability of the dependent variable which can be accounted for by the independent variable(s) in a regression model. MAE (Mean Absolute Error) is the statistic which is used to evaluate the performance of regression models. The MAE measures the average magnitude of errors in a set of predictions, without considering their direction (over or under the actual value). RMSE (Root Mean Squared Error) is an essential measures in regards to the performance of regression models. The RMSE computes the magnitude of deviations between the population or sample expected values and their corresponding actual values of a dataset.

### 3.5 System Deployment

Because of best performance LightGBM model was serialized and integrated into a Flask Based web application. Model Serialization is by trained LightGBM model is saved as a .pkl file for prediction on new data. Flask API development was added to accept user input as features in a web form, and return predicted order quantity in real time. For DataBase MongoDB Integration made for user name storage, user authentication, and session management. Finally User Interface was build by responsive HTML, CSS, and JavaScript for front-end was made in order to make it easy for food preparation management and staffing to use.

### 3.6 Proposed System Workflow

Fig. 1 show the work flow of the planned system the first data accumulation phase of the proposed food order prediction systems work flow starts with the collection of previous orders and contextual data including day of week, food type, temperature, weather, discounts available and holiday status. After that the dataset is loaded and preprocessed, separating it into training and testing sets, dealing with missing values and encoding categorical variables. Various machine learning models, such as, XGBoost, LightGBM, RandomForest, Gradient Boosting and ElasticNet, are initialized and trained on the preprocessed dataset during the model training phase. Next, we evaluate the models use good performance metrics including RMSE, MAE and R2 score, because to establish which model performs best. After a best model is determined the system can assist restaurants and catering organisations to forecast demand by allowing up to date input parameters by the customers via a web interface and provide predicted food orders in real time.

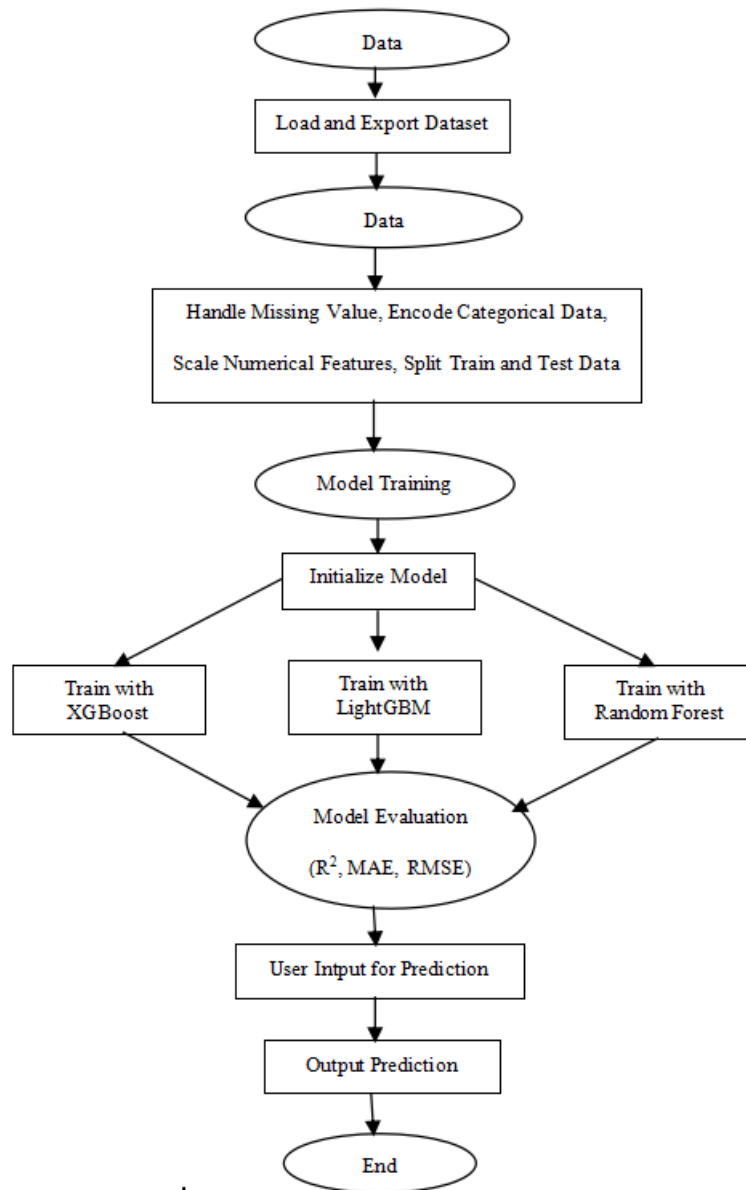


Fig 1: Proposed System Workflow

#### IV. NUMERICAL ANALYSIS

##### 4.1 Performance Metrics

The three primary evaluation components used for model evaluation are Mean Absolute Error which is the average absolute distance between food orders predicted and food orders made, lower is better. Root Mean Squared Error which is a quadratic measure of prediction error that penalizes large deviations more than the MAE. R<sup>2</sup> Score (Coefficient of Determination) The percentage of variance in food orders that the model can account for Numbers close to 1 are preferable.

##### 4.2 Model Performance Results

The models was trained on 80% of the dataset and then tested on the remaining 20% dataset. Table 1 illustrate the numerical results obtained.

Table 1: Comparison Of Models

Model	MAE	RMSE	R <sup>2</sup> Score
ElasticNet Regression	81.05	106.24	0.6675
Random Forest Regressor	57.06	81.52	0.8043
Gradient Boosting Regressor	52.27	73.83	0.8395
XGBoost Regressor	31.52	48.75	0.9300
LightGBM Regressor	31.66	46.82	0.9354

Table 1: shows performance metrics of machine learning models of food order prediction.

#### 4.3 Interpretation of Results

In the data, LightGBM Regressor has the least RMSE (46.82) and the highest R2 score (0.9354), which exceeds any of the other models. This presents that it is very proficient in capturing both continuous and categorical characteristics along with complex non-linear relationships in the dataset. In addition, XGBoost Regressor has shown great performance with the least MAE (31.52) and a R2 of 0.9300. XGBoost had a slight increase in RMSE (48.75) compared to LightGBM, however, it is still a very competitive and reliable model to forecast food orders. Gradient Boosting Regressor showed a R2 of 0.8395, MAE of 52.27, and RMSE of 73.83.

#### 4.4 Feature Importance

Feature importance is taken from the LightGBM model for understanding which factors most influence food order prediction. The below table shows the relative importance scores in percentage:

Table 2: Feature importance of food order prediction from LightGBM model.

Feature	Importance (%)
Last_Year_Orders	94
Temperature	50
Weather_Sunny	14
Delivery_Mode_Dine-in	12
Other Features	10

Table 2 Shows the last year's orders were previously the highest since they made up about half the prediction power of the model. The last year's total orders were expected to be the highest because temperatures have been shown to strongly influence demand for consumers. Contextual features like Weather\_Sunny and other delivery options like dine-in also are important dimensions of that class. When pairs with the reported results above, the two basically say that historical demand and contextual features are the two dimensions that support accuracy in order prediction.

#### 4.5 Error Analysis

Additional examination of the errors identified that the models exhibit larger prediction errors for mid-range order volumes, due to added variation in customer behavior on average-demand days. In contrast, the ML models are better suited to predict customer demand when there are either high-demand peaks (e.g., weekends, discounts) or low-demand periods (e.g., mornings, holidays), when the ordering behavior is more discernible.

V. RESULTS

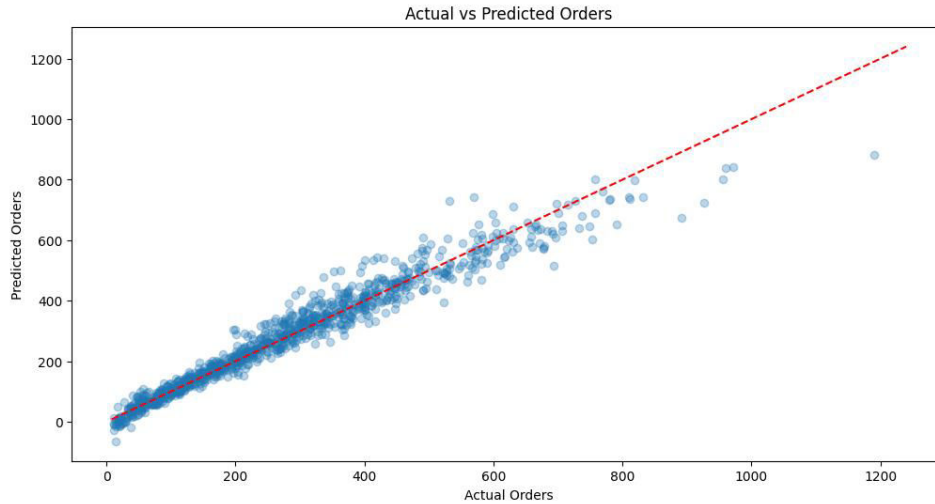


Fig 2: Actual Orders and Predicted Orders

Fig. 2 displays the comparison of actual and predicted food orders from the proposed machine learning model. The scatter plot identifies there are data points close to the red dashed line, indicating high accuracy in predictions. These are data points near the dashed marked line indicates tight clustering along the red dashed line and small error between actual and predicted. There is a very strong correlation between the actual value and the predicted value.

Model Performance Comparison

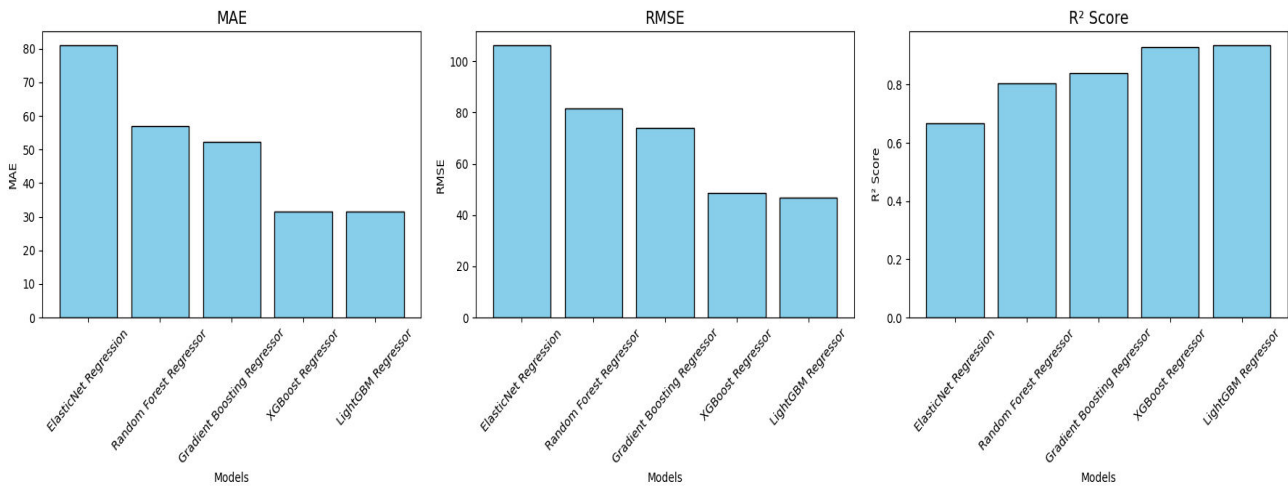


Fig 3: Model Performance Comparison

Fig. 3 Each bar graph shows performance of a machine learning models we used in project for the given metric evaluation which are MAE, RMSE, R<sup>2</sup>. X axis shows Machine Learning models, ElasticNet Regression, RandomForest Regressor, GradientBoosting Regressor, XGBoost Regressor, LightGBM Regressor. Y axis Metric values.

- A. Left graph: Mean Absolute Error (MAE).
- B. Middle graph: Root Mean Squared Error (RMSE).
- C. Right graph: R<sup>2</sup> Score.

The relative model performance of various models across MAE, RMSE and R2 metrics is shown in the bar chart comparison. LightGBM and XGBoost have the least mistakes, and also have the least values on the recommended lower end scales of MAE and RMSE. ElasticNet has the largest error figures; thus, it has the least level of prediction power. In terms of providing the greatest R2 score (~0.935), LightGBM takes the top spot, with XGBoost following closely behind (~0.93) and both manage to capture much of the data variance. ElasticNet continues to remain the worst across all metrics though Gradient Boosting did quite well.

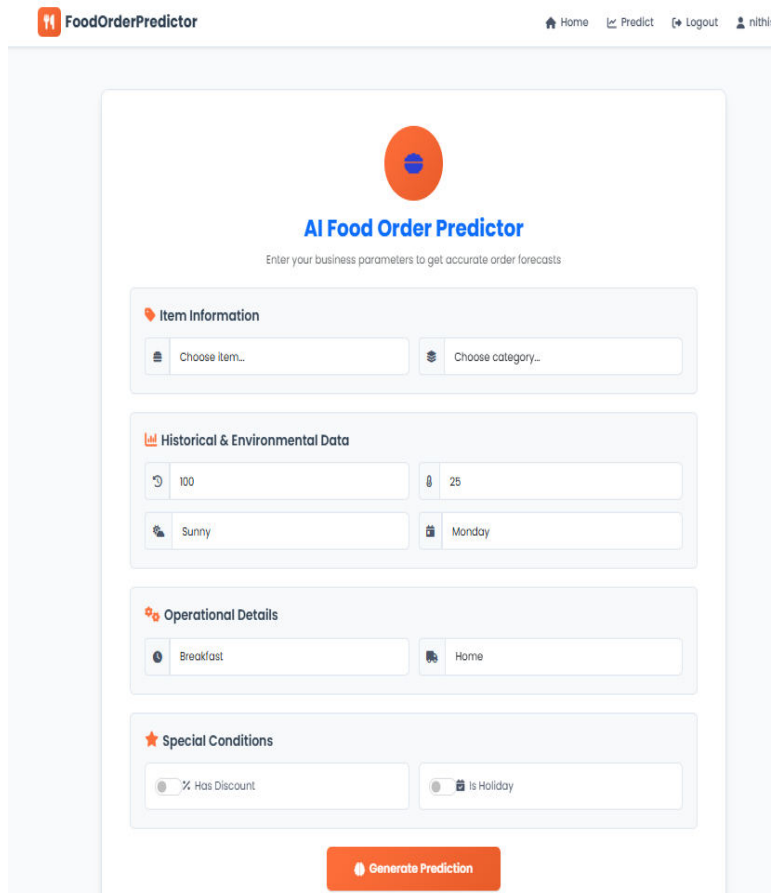


Fig 6: User Interface

Fig. 6 is the user interface of prediction of food orders. User After filling all the parameters, we click Generate prediction button. Once that Generate Prediction button was pressed, the model predicts the output as how much orders for that day of specific item.

## VI. CONCLUSION

This work can predict food order requests by identifying several factors such as day of the week, weather, presence of discount, holidays, and delivery method. This work employed 5 different algorithms such as XGBoost, LightGBM, Random Forest (RF), GradientBoosting, and ElasticNet. After testing the different models we realized that LightGBM get top prediction accuracy with low error rate, according to the results we generated this method could assist situated restaurants, catering services and delivery platforms with forecasting enough leads to predict demand in advance, planning efficient use-of-resources, and waste reduction. In Future using Long Short Term Memory (LSTM) and Recurrent Neural Network (RNN) might be high effective in identifying complex temporal relationships. Factors like social events, local festivals and regional public holidays may improve impact on people eating food. Deployment of Cloud in future may reduce issues in scalability, accessibility.

**REFERENCES**

1. R. S. Mallik, R. Abhiram, S. R. Reddy, and R. M. Jagadish, "A Comprehensive Survey on Sales Forecasting Models Using Machine Learning Algorithms", 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2022.
2. S. M. N. Sakib, "Restaurant Sales Prediction Using Machine Learning", International Research Journal of Engineering And Technology (IRJET), Volume 10, 2023.
3. Mehmet. ACI and D. YERGÖK, "Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory", ISSN 1330-3651, Volume 30, 2023.
4. V. B. Renukaprasad, "Decision Support System For a Restaurant to Forecast Sales using Machine Learning Techniques", M.Sc. dissertation, Data Analytics, Dublin Business School, Dublin, Ireland, 2020.
5. B. Madani and H. Alshraideh, "Predicting consumer purchasing decisions in the online food delivery industry", Computer Science & Information Technology (CS & IT), 2021.
6. H. M. Merdas and A. H. Mousa, "Food sales prediction model using machine learning techniques", International Journal of Electrical and Computer Engineering (IJECE), Volume 13, Dec, 2023.
7. N. Hübner, J. Caspers, V. C. Coroamă, and M. Finkbeiner, "Machine-learning-based demand forecasting against food waste: Life cycle environmental impacts and benefits of a bakery case study", Institute of Environmental Technology, Technische Universität Berlin, Berlin, Germany, Volume 28, 2024.
8. K. Posch, C. Truden, P. Hungerländer, and J. Pilz, "A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants", International Journal of Forecasting, Volume 38, 2021.
9. T. Tanizaki, T. Hoshino, T. Shimmura, and T. Takenaka, "Restaurants store management based on demand forecasting", Conference on Intelligent Computation in Manufacturing Engineering, Volume 88, 2020.
10. A. S. Bozkir and E. A. Sezer, "ADEM: An online decision tree based menu demand prediction tool for food courts", International Conference on Food Engineering and Biotechnology (IPCBE), Volume 50, 2013.
11. K. Lutostawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk, and A. Kozina, "Food demand prediction using the Nonlinear Autoregressive Exogenous Neural Network", IEEE Access, Volume XX, 2021.
12. Akshay Krishna, Akhilesh V, Animikh Aich, Chetana Hegde, "Sales forecasting of Retail Stores using Machine Learning Techniques", (IEEE)International Conference on Computational Systems and Information Technology for Sustainable Solutions, IEEE Xplore, 2019.
13. Abhinesh P. V., Ohmprakash S., and Prakash B., "Food Sales Analysis and Prediction Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume5, 2023.
14. A. Vaidhy, D. Batham, R. Jain, and A. K. Manjhwari, "Machine Learning-Driven Statistical Analysis of Indian Restaurants: Insights from the Zomato Dataset", Prestige Institute of Management and Research, Volume 38, 2025.
15. N. Nassibi, H. Fasihuddin, and L. Hsairi, "Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches", International Journal of Advanced Computer Science and Applications, Volume 14, 2023.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# International Journal of Multidisciplinary and Scientific Emerging Research (IJMSERH)

**Impact Factor: 9.274**