IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

FOOD DEMAND PREDICTION USING MACHINE LEARNING ALGORITHMS

Mr.Vijayanarayanan.A[1], Arun kumar G [2], Dhana Sehwac [3], Immanuel B[4],
[1]AssistantProfessor,[2,3,4] Students,
Department of Computer Science and Engineering,
PERI Institute of Technology, Chennai, India.

***Abstract**:* Population expansion, climate change, and digitization are driving up food demand faster than the country's economic growth. Perishable goods are handled more frequently by the food-related businesses, such as fast food chains, restaurants, canteens, and catering services. Organizing consumer food orders is one of the main problems faced by the food-related industry. Occasionally, inaccurate ordering estimates might result in either too much or too little food, wasting both food and raw materials and cutting into profits for the business. In order to ensure sustainable food systems and efficient resource allocation, food demand forecast is crucial. In an era of rapidly rising populations and shifting consumer tastes, accurate food demand forecasting is critical to minimizing food waste, optimizing performance, and enhancing food security. In this quest, machine learning algorithms have proven to be effective tools, with the ability to produce forecasts that are more dynamic and exact. In order to analyze the historical data and forecast the consumption for the upcoming months, we employed the decision tree algorithm. Along with internal variables, machine learning models also incorporate external variables like social events and economic indices.

Keywords—Population growth, Food-related Industry, Resource allocation, Food waste, Machine learning algorithms, External variables

I. INTRODUCTION

Demand forecasting is a vital tool that offers vital information to many different sectors. By 2050, there will likely be more people on the planet than 9.7 billion, which will contribute to global hunger and food insecurity. Accurate demand forecasting is necessary to direct organizational strategy planning and decision-making procedures. Forecast accuracy and reliability are critical because they play a key role in influencing business decisions. Demand forecasting uses appropriate forecasting methodologies in conjunction with past demand data to anticipate future demand. Food demand forecasting makes sure that the proper quantity of food is produced, shipped, and distributed to satisfy customer demands. Hotels and catering services can enhance customer satisfaction by tailoring their offerings based on predicted demand. The aim of food demand prediction is to guarantee a varied and visually appealing menu that corresponds with the expectations of the patrons. It encourages sustainable farming and food production methods, which lessen their negative effects on the environment and the wise use of resources. The demand for different foods may be predicted, which helps hotel operators control inventories and cut costs.

2 MATERIALS AND METHODOLOGY

2.1 Data Set : The dataset used in the study consists of one-year daily menus and corresponding sales numbers obtained from various hotels. The dataset covers diverse features such as weekdays, main courses, side dishes, soups, and the number of individuals, among others. Through training these features using different combinations and methodologies, the aim is to attain the highest level of accuracy in estimating the number of individuals expected to dine on any given day.

2.2 Preprocessing of data : Preparing data for modeling requires a series of critical steps, including managing missing values, cleansing data, scaling features, encoding categorical variables, feature selection, dimensionality reduction, normalization, data transformation, handling imbalanced data, and preprocessing text. Accurate model training and prediction are made possible by these techniques, which guarantee that the data is high-quality and presented correctly.

.3 Dataset Partitioning : The data set is partitioned when it is split into two sections, training and testing. By using the training set, the model is trained and made able to recognize patterns and correlations in the data. The testing set is used in the interim to assess the model's performance on data that it did not come into contact with during training.

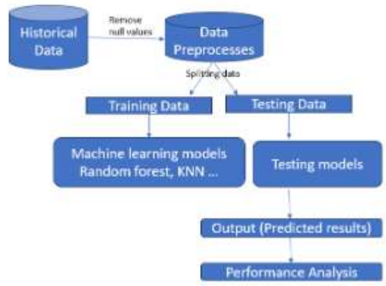


Figure 1 System Architecture

2.4 Machine Learning Algorithms:

2.4.1 Random Forest : Random Forest is a robust machine learning algorithm applicable to food demand prediction. This process involves forecasting the quantity of food needed based on various factors. Initially, relevant data is collected, encompassing historical demand, seasonal patterns, economic indicators, and demographic information. Subsequently, the data is pre-processed, handling missing values and normalizing features. The dataset is subsequently split into training and testing subsets. Continuous monitoring and periodic updates are essential to ensure the model remains accurate and relevant. The Random Forest algorithm's strength lies in its versatility, accommodating both regression and classification tasks. Overall, Random Forest emerges as a powerful tool in addressing the complexity of food demand forecasting, leveraging its ability to handle diverse data and deliver accurate predictions.

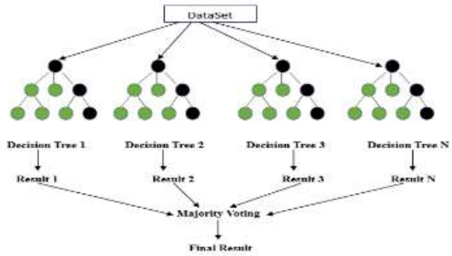


Figure 2 Random Forest Algorithm

2.4.2 XG Boost Algorithm : XG Boost, known as "Extreme Gradient Boosting," is a meticulously optimized distributed gradient boosting library specifically engineered for efficient and scalable machine learning model training. Widely adopted for its capacity to manage extensive datasets, XG Boost leverages machine learning algorithms efficiently. Notably, one of its prominent features is its adept handling of missing values, minimizing the need for extensive preprocessing of real-world data. Moreover, XG Boost incorporates native support for parallel processing, enabling the training of models on large datasets within reasonable timeframes. Its methodology involves iteratively constructing a multitude of decision trees, each subsequent tree aiming to rectify the errors of the ensemble formed by its predecessors. This iterative process continues until an optimal model is attained.

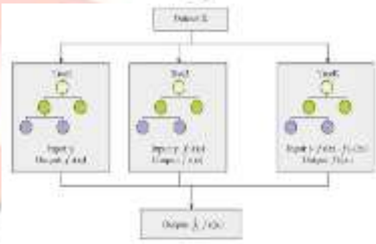


Figure 3 XG Boost Algorithm

2.4.3 Light GBM : Light GBM is a gradient-boosting framework based on decision trees to increase the efficiency of the model and reduce memory usage. Light GBM stands as a robust and resource-efficient gradient boosting framework, open-source and tailored for machine learning tasks. Light GBM is tailored for efficient processing of extensive datasets, prioritizing speed and memory utilization. Employing gradient boosting, it merges numerous weak learners, typically decision trees, to forge a potent predictive framework. Light GBM grows trees in a leaf-wise manner, focusing on the nodes that contribute the most to the reduction in the loss function. This approach often leads to faster convergence and more efficient use of computational resources.

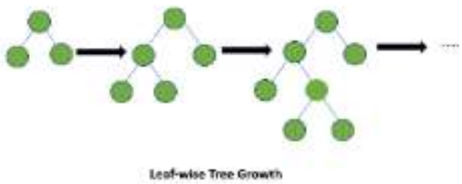


Figure 4 Light GBM Algorithm

2.4.4 Lasso Regression : Lasso, an acronym for Least Absolute Shrinkage and Selection Operator, is widely used in machine learning to handle high-dimensional data through automated feature selection. This is accomplished by introducing a penalty term to the residual sum of squares (RSS), which is then adjusted by the regularization parameter (lambda or λ). The regularization parameter governs the extent of regularization applied. Higher lambda values increase the penalty, encouraging more coefficients to approach zero, thus reducing the importance of specific features or even eliminating them entirely from the model, thereby enabling automatic feature selection. Conversely, smaller lambda values mitigate the penalty's impact, allowing more features to be retained within the model.

2.4.5 Elastic Net Regression : Elastic net linear regression integrates penalties from both lasso and ridge techniques to regulate regression models effectively. By leveraging the strengths of both methods while addressing their weaknesses, elastic net enhances the regularization of statistical models. Unlike lasso, which struggles with high-dimensional data by selecting only a few samples, elastic net allows the inclusion of numerous variables until reaching a saturation point. Furthermore, lasso tends to select only one variable from highly correlated groups, disregarding the others entirely. To overcome these limitations, elastic net introduces a quadratic expression (||β||2) in the penalty, reminiscent of ridge regression when used independently. This quadratic expression promotes convexity in the loss function, combining the advantages of lasso and ridge regression.

2.4.6 K-Nearest Neighbours Regression : KNN regression, a non-parametric approach, provides an intuitive means of estimating the relationship between independent variables and continuous outcomes by averaging observations within the same neighborhood. The analyst must specify the neighbourhood's size or utilize cross-validation to determine the size that minimizes mean-squared error. Despite its attractiveness, the method becomes impractical as the dimensionality increases, particularly with numerous independent variables.

2.4.7 Gradient Boosting Regressor : Gradient boosting, an ensemble machine learning technique, employs weak learners sequentially to build a robust model. This versatile and potent method is applicable to both regression and classification tasks, often yielding favourable results with minimal tuning. It demonstrates proficiency in handling a vast number of features without bias toward any specific feature type. However, gradient boosting is more prone to over fitting compared to other machine learning approaches and may exhibit slower training speeds, particularly with large datasets. Nonetheless, despite these drawbacks, gradient boosting remains a favoured choice for numerous machine learning tasks, owing to its flexibility, efficacy, and overall good performance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

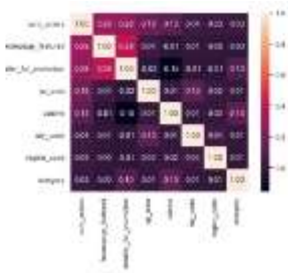
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

2.5 Evaluation Metrics: Evaluation techniques and metrics used for evaluating food demand prediction models: • Mean Absolute Error (MAE): Computes the average absolute difference between predicted and actual values. • Mean Squared Error (MSE): Calculates the average squared difference between predicted and actual values. • Root Mean Squared Error (RMSE): The square root of MSE, indicating error in the same unit as the target variable. R-squared (R2): Assesses the proportion of variance in the dependent variable explained by the model.

Where N represents the total number of observations, ŷ denotes the predicted value of y, and y signifies the mean value of y.

4. Results and Discussions : • The amount of orders is positively correlated with the highlighted webpage and promotional mailing.



• The quantity of orders is directly correlated with the area and cuisine. There is a negative association between the cuisine and the area when it comes to the webpage and promotional mailing. Many characteristics have no link at all and may be independent. Using feature engineering, it is possible to combine a number of unrelated features. Feature engineering essentially creates a suitable dataset, which facilitates and eases data analysis. It also makes the training model perform better. Labeled coding allows them to categorize the categorical data without losing any important information and converts them into a numeric format. It is also seen that with Bayesian Linear Regression ,KNN ,Light BGM, LASSO have very low accuracy with respect to XG Boost and Gradient Boosting Regressor. Decision Tree algorithm has given the highest accuracy and thus gives us better performance with respect to other models.

5. CONCLUSION

In conclusion, food demand prediction using decision trees is a valuable and practical approach. Decision trees offer a transparent and interpretable way to model and understand the complex factors influencing food demand. While traditional statistical methods such as linear regression have played a foundational role in food demand prediction, they come with certain limitations in handling the complexity and non-linearity of real-world scenarios. As consumer preferences change and external factors influence, more advanced machine learning techniques are increasingly embraced to deliver more accurate and adaptable predictions.

REFERENCES

[1] Time Series Forecasting and Modelling of Food Demand Suplly based on Regressors Analysis ,IEEE (2018)

[2] Demand Prediction model for forecasting AGRI-needs of the society, International Conference on incentive computing and Informatics (2016)

[3] Food Demand Prediction using Statistical and Machine Learning models , Master Thesis National College of Ireland (2023)

[4] Predicting Long Term Food Demad, Cropland use, and Prices, Annual Review of Resource Economics (2016)

[5] Patrick Meulstee amd Mykola Pechenizkiy, Food Sales Prediction, International Conference on Data Mining (2018)

[6] A Review of Food Demand Forecasting Models and Machine Learning Techniques by Wei Zheng and Xuemin Zhang (2019)

[7] Deep Learning for Food Demand Forecasting: A Review and Future Directions by Yuhao Kang, Zhiyong Chen, and Liyuan Liu (2020)

[8] Predicting Food Consumption Using Machine Learning Techniques: A Systematic Literature Review by Fatemeh Ramezanpour and Yudho Giri Sucahyo (2020)

[9] A Hedonic Price Model of a Demand for Breakfast Cereals and the Impact of Promotional Activity" by Darius J. Lakdawalla and George J. Zanjani (2005)

[10] M.A. Yurdusev, M. Firat, M. E. Turan, Generalized regression neural networks for municipal water consumption prediction, Journal of Statistical Computation and Simulation (2010).

[11] Patrick Meulstee and Mykola Pechenizkiy, Food Sales Prediction2008 IEEE International Conference on Data Mining Workshop.

[12] Comparison of Machine Learning Techniques for Food Demand Forecasting in the Fast Food Industry by Haidar Akmar, Muhammad Adib Surani, and Muhammad Hanif Abd Rashid, published in the International Journal of Recent Technology and Engineering in (2020).

[13] A Novel Forecasting Model for Food Demand Based on ARIMA and LSTM Neural Network by Yanying Shao, Yafei Zhang, and Yuzhou Wu, published in the IEEE Access journal in 2020

[14] Demand Forecasting for Fresh Food Retailers: A Comparison of Machine Learning Methods by Yuwen Chen, Yuqian Xu, and Xi Chen, published in the International Journal of Production Economics in (2021).