iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50283

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Anticipating Consumer Demand using ML

P. Rama Devi[1], Srujitha Meesala[2], Ramya Reddy[3], Kushal Senapathi[4], Udaya Kolala[5]
[1]*Assistant Professor,* [2,3,4,5]*Students, Department of Computer Science and Engineering, Gandhi Institute of Technology and Management (GITAM), Visakhapatnam, Andhra Pradesh, India.*

Abstract: ***Demand forecasting is essential for every growing online business. Without efficient demand forecasting systems in place, it might be next to impossible to always have the right amount of stock on hand. Because a food delivery service deals with a high volume of perishable raw materials, it is critical for the company to accurately forecast daily and weekly demand. If a warehouse has too much inventory, there is a greater likelihood of wastage, and if it has too little, there may be shortages, which would encourage customers to turn to your competitors. Therefore, predicting demand is one of the important tasks to be done. The project represents a food delivery company that operates in multiple cities. This particular company has various fulfillment centres in these cities for dispatching meal orders to their customers. Its objective is to anticipate consumer demand and the goal is to build a predictive regression model to assist the client in projecting demand for the following weeks so that these centres can organize their raw material stock properly, with the usage of various Machine Learning and Deep Learning Models and Techniques. For that purpose, there are various tools, techniques and methods are proposed. Linear regression model, Random Forest, XG Boosting, Decision tree is some of the models performed for getting the highest accuracy.***
Keywords: Forecasting, Food Inventory, Deep Learning, Regression models.

I. INTRODUCTION

A forecast is a projection of what will happen in the future, specially concerning a specific situation or event. Demand forecasting is a mechanism that impacts our understanding of the variables that influence the variable we desire to forecast. Demand forecasting is the practice of estimating a product or service's future demand based on historical data, current market conditions, and other relevant variables. Demand forecasting is used to assist organizations in planning their production, inventory, and staffing levels and to make sure they can timely and economically meet customer demand. Demand forecasting can be done using various methods, such as time series analysis, regression analysis, and models based on artificial intelligence or machine learning. The choice of approach relies on the nature of the business, the accessibility of data, and the needed level of accuracy. Each technique has pros and limitations of its own. Hence, to enhance the output and maintain the integrity of the event, Various algorithms are proposed in ML and deep learning models to predict demand.

Nowadays, managing Inventory is actually a prevalent issue. In order to maintain the consumers and prevent food shortages, inventory management is crucial. It's crucial to stock the shelves and maintain track of the inventory. Providing for clients' eating needs completely satisfies them. Companies may determine how much goods to order and when by using inventory management. Making sure there are enough products or materials to meet demand while avoiding overproduction or surplus inventory is the basic goal of inventory management.

The advantages of predicting inventory include ensuring and maintaining good customer service, facilitating the flow of items through the production process, achieving a reasonable utilisation of personnel and resources, and offering protection against supply and demand risks. The inventory managers will benefit from using this technology to monitor sales and determine whether certain items are still in stock. Therefore, this would assist employers in replenishing inventory as needed to prevent stock backlogs.

II. LITERATURE REVIEW

Time series forecasting is becoming more and more popular because to machine learning (ML) approaches. Recent studies have found that the volume of training data has an impact on the results. Using data from various time series, cross-sectional forecasting trains the ML model. It is sometimes referred to as cross-sectional training and was recently created to address the lack of data offered by short time series. The application of cross-sectional forecasting to supply chain demand will be examined in this thesis.

The project will be broken up into three sections, each of which will address one of the following topics: (i) applying cross-sectional forecasting to the entire dataset using a variety of ML methods; (ii) experimenting with four cross-sectional forecasting clustering approaches; and (iii) including exogenous variables in addition to historical demand data for forecasting using cross-sectional training.

©IJRASET: All Rights are Reserved

The tests, which included data from two datasets on food distribution, showed that the ML techniques surpassed the statistical benchmarks. Additionally, it was shown that the effectiveness of ML approaches could be rationally enhanced by employing the right clustering methodology and that they were able to account for the effects of extra demand-influencing variables, hence reducing forecasting inaccuracy.

III. PROBLEM IDENTIFICATION

A meal delivery service that has operations in several cities is your client. For delivering meal orders to clients, they have a number of fulfilment sites in these cities. The client requests your assistance in projecting demand for the following weeks so that these centres can organize their raw material supplies properly.

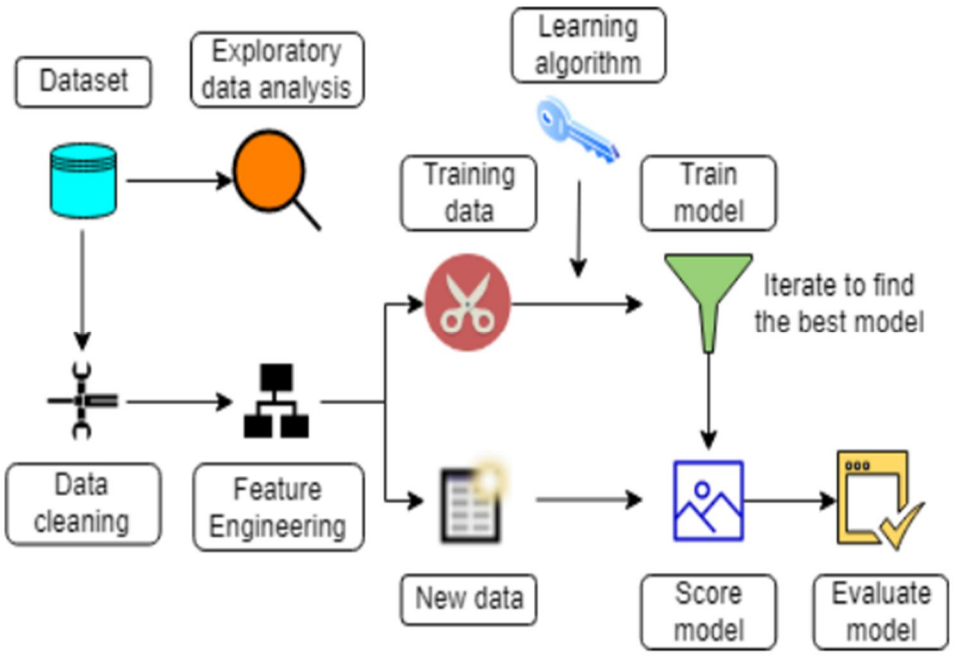
The overwhelming bulk of raw goods are replenished weekly, and because they are perishable, careful purchase planning is crucial. Second, precise demand estimates are quite beneficial when it comes to staffing the centres.

The aim is to forecast the demand for the centre-product combinations in the testing set for the upcoming weeks based on the following information:

1) Demand history for a meal-centre combination (145 weeks)
2) Features of the product (meal), including classification, subcategory, price, and discounts.
3) Details about the fulfilment centre, including its location and other details about the city.

IV. SYSTEM METHODOLOGY

The figure below shows how the dataset is cleansed and exploratory data analysis is performed to describe their primary characteristics. The cleansed data is then processed through feature engineering to extract the best features from the raw data and improve the performance of the machine learning model, and it is divided into training data and test data. The training data is then sent to a random forest model, which operates by creating a large number of decision trees during training. The test data is then fed into the model, which compares projected and actual results using an assessment metric called RMSLE. This model can then be used to forecast future sales data.



V. DATASET

The datasets considered here are; one is the *Weekly demand data* which contains the historical demand data for all the centre's; the *fulfillment centre info* which contains information for each fulfillment centre, and *meal info,* which contains information for each meal being served.

The available dataset contains 3 data tables labelled as follows:

1) Weekly Demand Data

Variable	Definition
id	Unique ID
week	Week No
center_id	Unique ID for fulfillment center
meal_id	Unique ID for Meal
checkout_price	Final price including discount, taxes & delivery charges
base_price	Base price of the meal
emailer_for_promotion	Emailer sent for promotion of meal
homepage_featured	Meal featured at homepage
num_orders	(Target) Orders Count

2) Fulfillment centre info.

Variable	Definition
center_id	Unique ID for fulfillment center
city_code	Unique code for city
region_code	Unique code for region
center_type	Anonymized center type
op_area	Area of operation (in km^2)

3) Meal info.

Variable	Definition
meal_id	Unique ID for the meal
category	Type of meal (beverages/snacks/soups....)
cuisine	Meal cuisine (Indian/Italian/...)

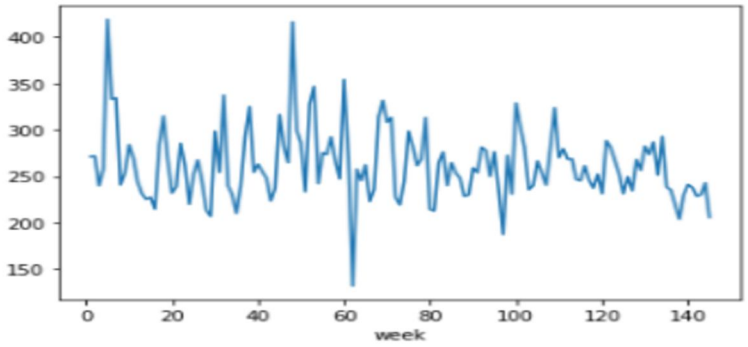
VI. DATA PRE-PROCESSING

Data pre-processing aids in the transformation of raw data into usable forms. To convert raw data into useable data, we used data cleaning techniques, exploratory data analysis, and feature engineering methods in this research.

A. Exploratory Data Analysis

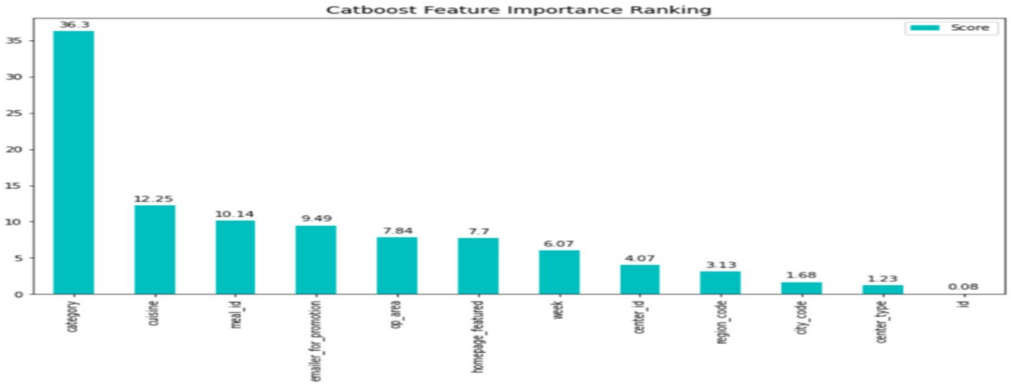
After analysing the data, it was determined that the number of orders placed (the target variable) is substantially right skewed, requiring the use of the Log transformation.

Visualizing the data-



B. Feature Selection

Feature Selection is a technique for limiting the input variable to your model by using only relevant data and removing noise from the data. It is the technique of selecting suitable characteristics for your machine learning model automatically based on the sort of problem you are attempting to solve.



VII. ALGORITHMS

The ML models used/considered for this project are:

1) Gradient Boosting
2) Linear Regression
3) Decision Tree
4) Random Forest

The Machine Learning forecasting approach entails:

a) Time Series Analysis
b) Regression Modelling
c) Deep Learning Modelling

Given our problem statement, we consider Regression Modelling to predict future outcomes. Theoretically, both XGBoost and Random Forest have the most accuracy out of the four models we have considered.

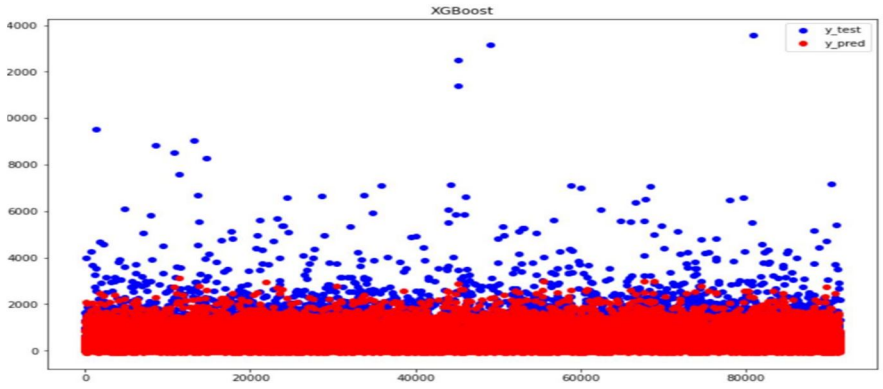
Therefore, we first process the data and then test the data through all these models and then consider the model with the most accuracy. Accuracy can be calculated or considered by finding MSE and R2 scores for each model. The model with lowest MSE (Mean Squared Error) or highest R2 score is considered to be most accurate.

VIII. RESULTS & ANALYSIS

The four algorithms are tested and evaluated

1) Gradient Boosting

R2 score: 0.5939403
MSE: 62771.911804

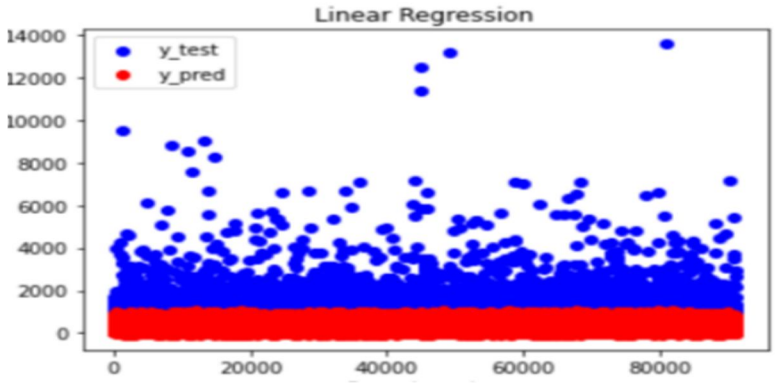


©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

2) Linear Regression

R2 score: 0.17266752

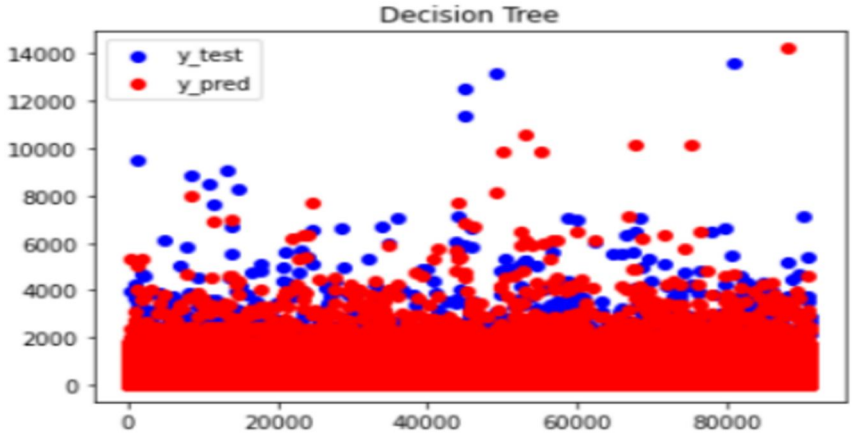
MSE: 127895.5990214



3) Decision Tree

R2 score: 0.757902931

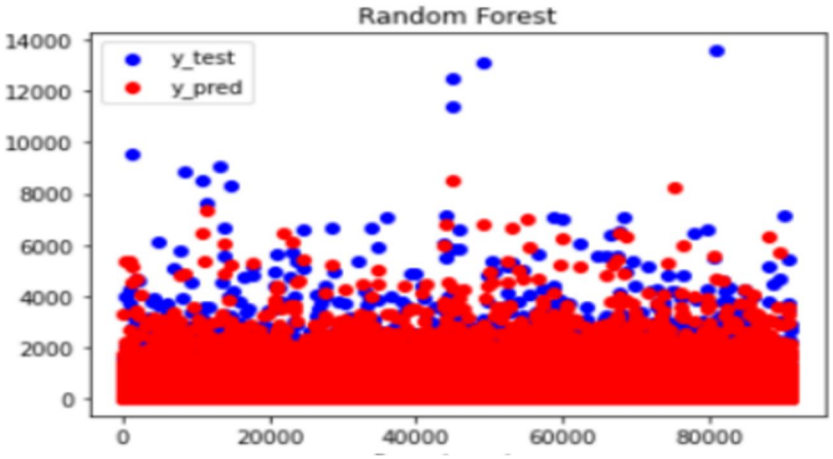
MSE: 37425.28010075



4) Random Forest

R2 score: 0.84833636

MSE: 23445.3644008

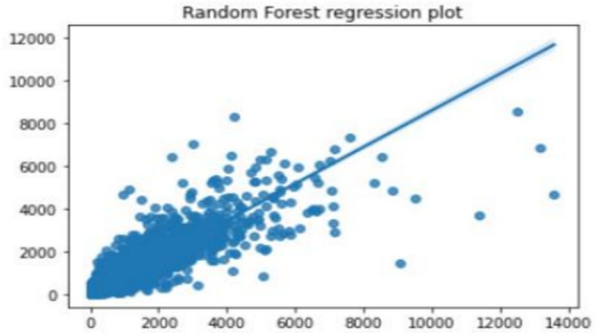


©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Summarizing the Findings and Analysis

R2 score (Coefficient of determination) is best in random forest compare to other models. To calculate the number of orders we will be using random forest for the above dataset.

The number of orders for the next 10 weeks are predicted using Random Forest Regression model.




```
final =  test.loc[:,['id','week','num_orders']]
final.head()
```

	id	week	num_orders
0	1028232	146	140.388137
1	1127204	146	172.763792
2	1212707	146	259.070000
3	1082698	146	114.278059
4	1400926	146	117.864612

```
                id  week  num_orders
0          1028232   146  140.388137
1          1127204   146  172.763792
2          1212707   146  259.070000
3          1082698   146  114.278059
4          1400926   146  117.864612
...            ...   ...         ...
32568      1250239   155   91.330000
32569      1039516   155   56.369762
32570      1158107   155  260.837500
32571      1444235   155  253.934333
32572      1291286   155  217.050000

[32573 rows x 3 columns]
```

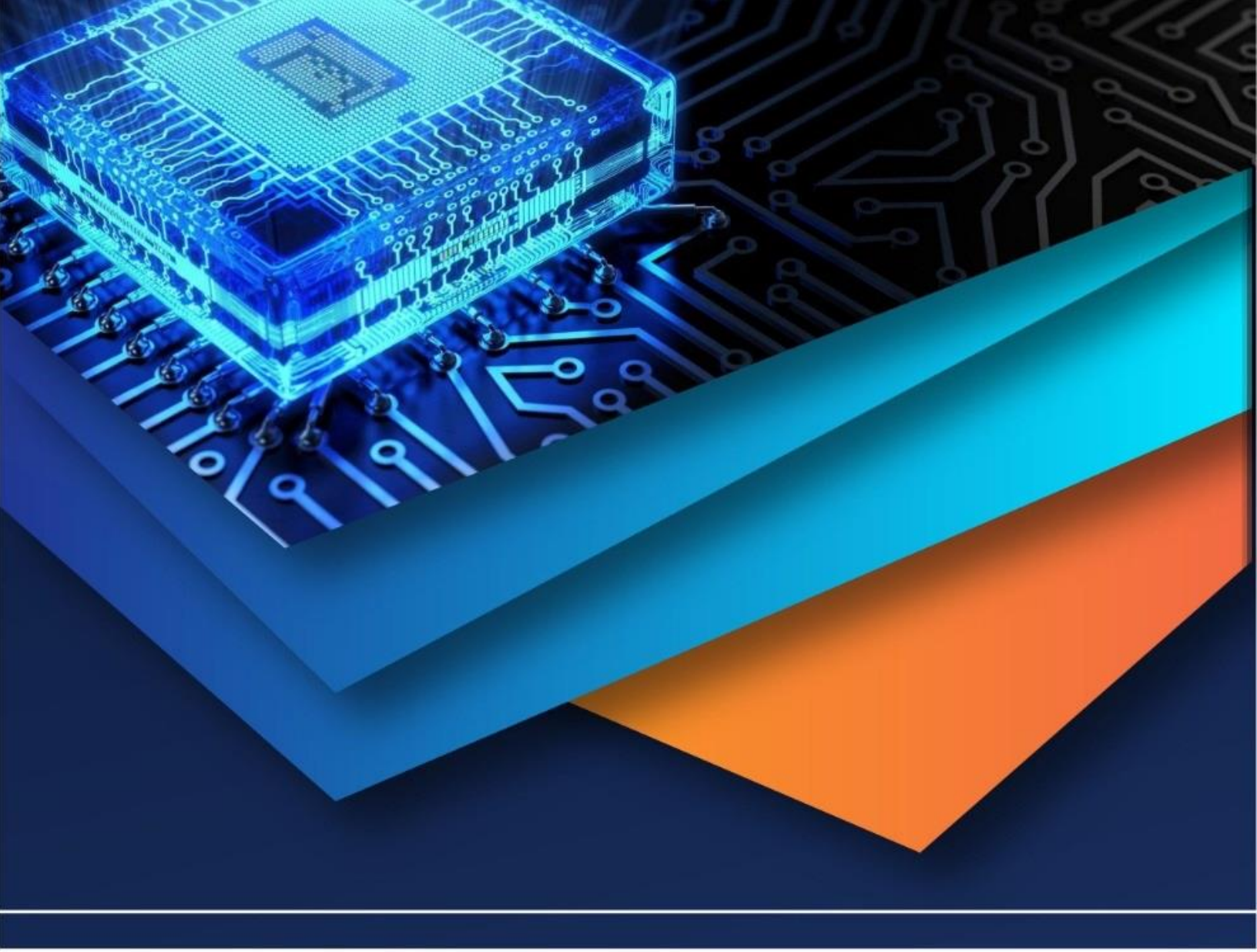
IX. CONCLUSION & FUTURE SCOPE

For the prediction in this study, which includes several criteria like area ID, week, etc., we are using both external and internal data. Food demand forecasting is a crucial and difficult task. problem. In this study, we discussed penalized regression, Bayesian linear regression, and the decision tree approach as food demand methods. The accuracy rate is rising as we use different algorithms for making predictions.

Furthermore, future predictions can be made with greater accuracy using a variety of other variables, such as cultural customs, religious holidays, consumer preferences, etc. In the future, this technique be utilized to foresee the need for a workforce and to automate the ordering of food.

REFERENCES

[1] Patrick Meulstee and Mykola Pechenizkiy, “Food Sales Prediction: If Only It Knew What We Know” 2008 IEEE International Conference on Data Mining Workshop.
[2] Yoichi Motomura, Baysian network, Technical Report of IEICE, Vol.103, No.285, pp.25-30, 2003.
[3] Yoichi Motomura, Baysian Network Softwares, Journal of the Japanese Society for Artificial Intelligence, Vol.17 No.5, pp.1-6, 2002.
[4] D. Adebanjo and R. Mann. Identifying problems in forecasting consumer demand within the fast-paced commodity sector. Benchmarking: An International Journal, 7(3):223– 230, 2000.
[5] Bohdan M. Pavlyshenko, “Machine-Learning Models for Sales Time Series Forecasting”, 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018.
[6] Irem Islek and Sule Gunduz Oguducu, “A Retail Demand Forecasting Model Based on Data Mining Techniques” – IEEE, October 2015
[7] https://statisticshowto.com/lasso/regression
[8] https://en.wikipedia.org/wiki/Random_forest
[9] https://towardsdatascience.com/supportvectormachines-svm-c9ef22815589
[10] https://towardsdatascience.com/httpsmedium-comvishalmorde-xgboost-algorithmlong-she-may-reinedd9f99be63



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)