International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

FOOD REQUIREMENT ANALYSIS IN AN AREA

[1] Latha K G, [2] Manjunatha H T

PG Student, Dept. of Computer Applications, Jawaharlal Nehru New College of Engineering, lathag790@gmail.com, Shivamogga, India

Assistant Professor, Dept. of Computer Applications, Jawaharlal Nehru New College of Engineering, manjudeepa@jnnce.ac.in, Shivamogga, India

ABSTRACT: - Demand forecasting is the process in which historical data is used to estimate the quantity of product customer will purchase. This prediction activity is used in many fields like retailing, food industry etc. In Restaurants, prediction play a vital role as most of the basic ingredients have short-shelf life. The demands depend upon many explicit and hidden context such as season, region etc. In this paper, number of order is used to forecast stock of items, using machine learning with internal and external data. In this we provide an appropriate algorithm for demand forecasting which is capable of overpowering the wastage of short life items. Proposed algorithm like Bayesian Linear Regression, LASSO, XG-Boost algorithm are used that considerably improves the forecasting performance.

KEYWORDS: Bayesian Linear Regression, LASSO, Random Forest.

I. INTRODUCTION

The success of a restaurant not only depends on taste, ambience but also on service. The most important part among the services is serving fresh food. In order to provide this, the restaurants need to prepare food daily, this requires buying some of fresh self-life food products every day. The major task that one would face in this will be predicting the quantity of products to be bought and prepared. It is very difficult to predict the number of orders in a given restaurant on a given day. A wrong prediction may end up purchasing and preparing less amount of food which will cause shortage or purchasing and preparing more which will lead to wastage of food. So, predicting the exact demand is a challenge because of uncertainty and fluctuations in consumer

demand.

These variations ad fluctuations in demand may be because of price change, promotions, changein customer's preferences and weather changes. All these factors imply that some dishes are sold mostly during limited period of time. Although we know that some regularseasonal pattern is expected, the features that predict these seasons are not directly observed. Thus, drops and rises inorders because of these seasonal changes are difficult to predict. In order to solve such problems, we are researchinghow to predict the demand. Here we are researching food demand forecasting methods using internal data such as number of orders.

II. RELATED WORK

Here we have selected few key literatures after exhaustive literature survey and listed as below:

Tanizaki et al. (2019) [1]-"Demand forecasting based on a number of customers" Describes: Since the accuracy of the boosting model is not significant, in comparison. The researcher employs boosting algorithms, XG-Boost, Light GBM, and cat-boost algorithms, to increase the forecasting rate. Among these boosting models, the XG-Boost model out preformed the previous research in this study.

Wibowo and Yasmina et al. (2021) [2]-"Forecasts actual food consumption demand" Describes: In order to assess the forecasting rates on the demand forecasting, many linear regression models, including lasso and ridge, multiple linear, and Bayesian ridge regression as well as non-linear regression models, such as SVR, Decision Tree, Random Forest, and Boosting models are utilized on a large dataset.

Bozkir and Sezer et al. (2011) [3]-"Forecasts actual food consumption demand" Describes: The researcher tired to build SVR model, but the model takes more time about more than 24 hours to get trained. However, the hyperparameters are not improved the models training time and the model was trained with 3 years of data including 423k observations.

Tanizaki et al. (2020) [4]-"Forecasting customer order quantity of beers" Describes: The models are trained using a sizable dataset containing 423k observations. All regression models have good average accuracy, which is over 80%. Hence the model's forecasting rate is good while predicting a number of food orders in a restaurant.

Posch et al. (2020) [5]- "Illustrates the demand prediction system by predicting future sales of daily sold quantity for the restaurants and canteens"Describes: The Facebook prophet model was built to accommodate

the details about the weather, special events, and holidays but the dataset contains weekly observations so I am unable to provide additional information while building the model. It would be better suited if it's a daily observation.

Menculini et al. (2021) [6]-"predictions of food sale prices for food goods" Describes: The researcher attempted to build Facebook prophet model on a weekly observation dataset. However, the accuracy of the model is comparatively low.

Fattat et al. (2018) [7]-"Demand Forecasting for food products in a food manufacturing industries" Describes: In order to assess the accuracy and forecasting rate, The researcher has built several statistical and machine learning models.

III. METHODOLOGY

In this research, the number of customers is forecasted using machine learning and statistical analysis method with internal data and external data in the ubiquitous environment. Bayesian Linear Regression, Boosted Decision Tree Regression, and Decision Forest Regression are used for machine learning, Stepwise method is used for statistical analysis method. We used Jupyter Notebook as a machine learning tool.

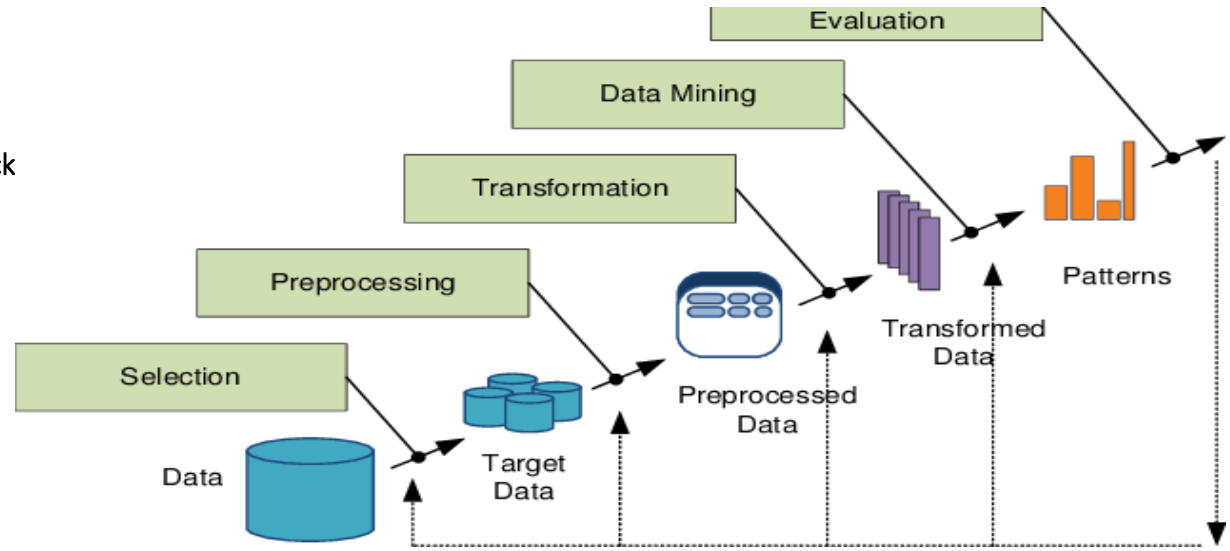


Figure 1: Block diagram

In order to extract useful information and patterns from a raw dataset, the researcher in this study used KDD (Knowledge Discovery from Databases) methodology. Fig.1 illustrates the stages of this method, which include goal setting, data selection, data cleaning and transformation, modeling, and evaluation.

I. Linear Regression

It is way technique which uses a Bayesian network for the aim of machine learning. We formulate linear regression using probability distributions instead of point estimates. The anticipated value of the variable is completed by the very best probability value of distribution of unobserved variables against observed variables. The conditional dependencies are often expressed in sort of a graph or data structure using this probabilistic model. It's mainly defined by three variables: conditional probability, variate variable and conditional dependency condition between random variables.

II. Bayesian Linear Regression

Bayesian Linear Regression (Bayesian) may be a method of applying Bayesian network to machine learning. The Bayesian network may be a probabilistic model during which conditional dependencies among multiple random variables are expressed employing a graph structure and dependency relationships between random variables are expressed by conditional probabilities. The Bayesian network is defined by three variables: variate variable, conditional dependency condition between random variables, and conditional probability. By using the Bayesian network, the probability distribution of unobserved variables is calculated using observed some variables and therefore the value with the very best probability value is obtained because the predicted value of variable.

III. Random Forest

It is a way which may be used for both classification and regression and deploys multiple decision tress to construct a forest and accumulates all the training results from each tree. It works well with both linear also as non -linear data, hence it relies on number of decision trees and uses mean prediction for the ultimate value .

IV. Support Vector Machine (SVM)

SVM may be a popular technique used for classification and builds a hyper plane to extract the data patterns. For SVM model to possess high accuracy, the training data must have top quality and relevant features, otherwise the performance would be very poor and would end in low accuracy [9]. Users can complete the

training tasks on non-linear distributions of coaching data by changing the kernel function of SVM.

V. LASS0

In statistics and machine learning, lasso could also be a multivariate analysis method that performs both variable selection and regularization so on reinforce, the prediction accuracy and interpretability of the statistical model it produces. Lasso regression could also be a kind of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, a bit like the mean[7]. The lasso procedure encourages simple, sparse models.This particular kind of regression is well-suited once wewould like to automate certain parts of model selection, likevariable selection/parameter elimination.

VI. XG-Boost

It is a recently new technique in field of machine learning which is predicted on the thought of gradient boosting. It uses decision tree method to supply high performance withvery less computation time, leading to better performanceswith real data. The number of incorrect predictions is increased to enhance the accuracy of the training model. Boost is an implementation of gradient boosted decisiontrees which is designed for speed and performance. Boost stands for extreme Gradient Boosting. It runs on one machine, also due to the distributed processing frameworks.It's many silent features such as-

- Clever penalization of treed.
- A proportional shrinking of leaf nodes.
- Newton Boosting.
- Extra randomization parameter

IV. FORECASTING OF NUMBER OF CUSTOMERS

On track variable

The meal delivery restaurant which is the client wants to forecast the orders for upcoming weeks. This is often a basic regression problem where model must predict the amount of orders for a specific week for a particular fulfilment center.For this, we can use extensive set of algorithms such as XG-Boost to solve the problem. Hence, if we observe weeklyand monthly trends, we find that-

- Week 5 and 48 had the highest number of orders, while week 62 has the lowest orders.
- It was also noticed that usually the first and last week of the month had highest number of orders as compared to other weeks.

Forecasting Method

- Numbers of orders have a slight positive correlation with homepage featured and mailer used for promotion.
- Number of orders also depends directly on cuisine and area.
- Area and cuisine have negative correlation with homepage it is featured in and mailer used for promotion.
- There are many features which have neutral relationship.

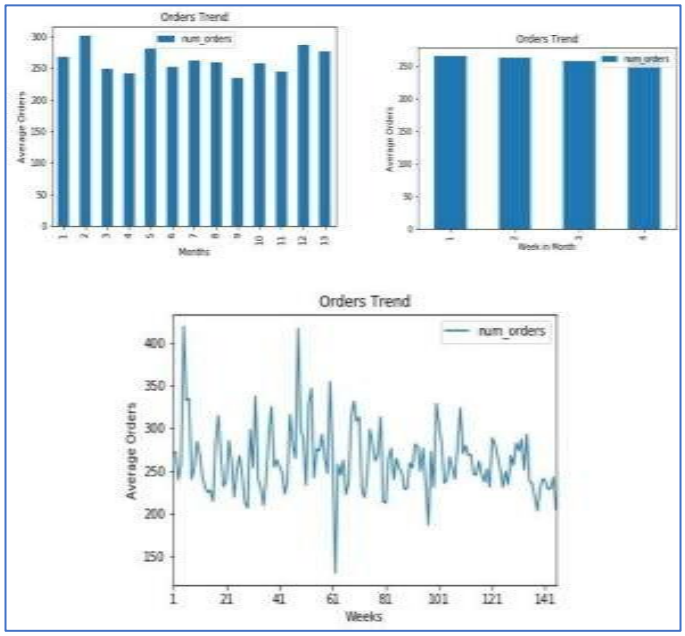


Figure-2 Showing orders trend according to week days

The figure-2 shows the following information, there are several irrelevant features which can be merged with the help of feature engineering. Feature engineering basically makes data analysis easier and compatible for analyzing by preparing a proper dataset. Also, it enhances the performance of the training model. Label encoding can be used for categorical data to convert them into numeric format and enables them to group the categorical data without losing any vital information. It is also seen that with Bayesian Linear Regression and LASSO have very low accuracy with respect to KNN and Decision tree. XG-Boost algorithm has given the highest accuracy and thus gives us better performance with respect to other models.

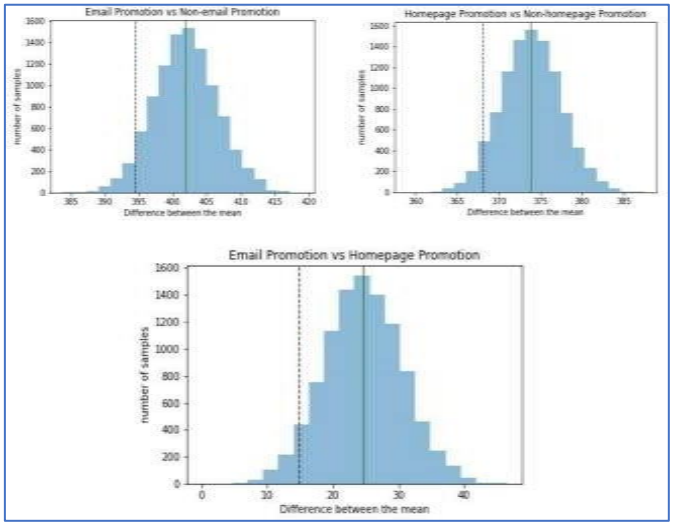


Figure-3(a)

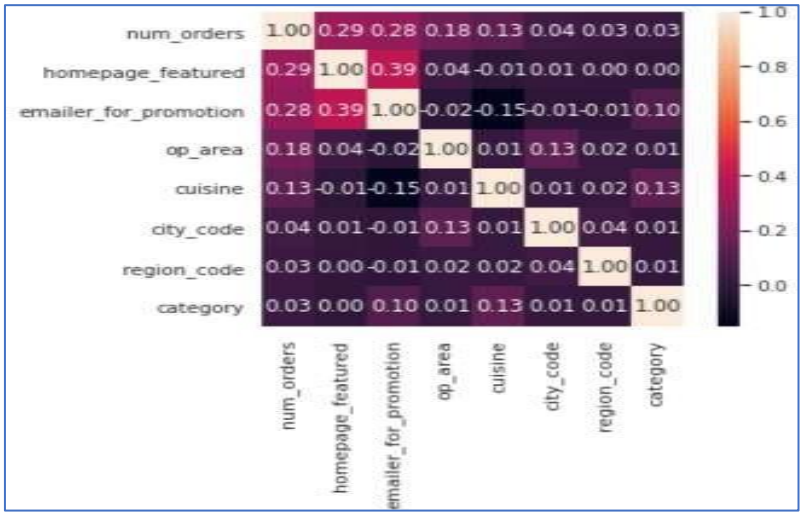
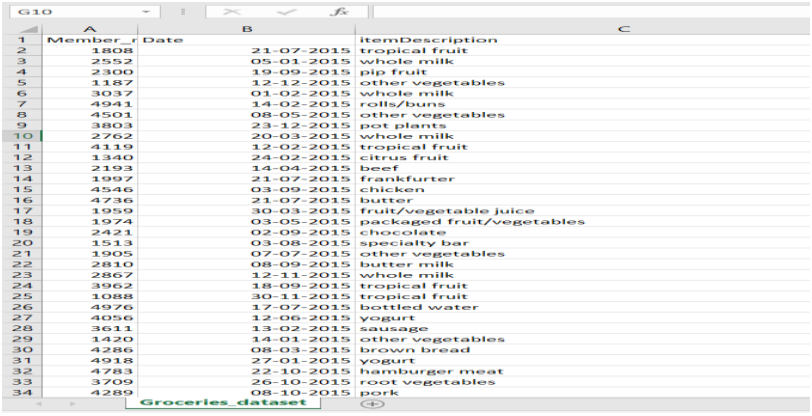


Figure-3(b)

Figure 3(a) shows the distribution of the data in the dataset file,

figure 3(b) shows correlation among various data items present in the data set

DATASET



	A	B	C
1	Member_r	Date	itemDescription
2	1808	21-07-2015	tropical fruit
3	2552	05-01-2015	whole milk
4	2300	19-09-2015	pip fruit
5	1187	12-12-2015	other vegetables
6	3037	01-02-2015	whole milk
7	4941	14-02-2015	rolls/buns
8	4501	08-05-2015	other vegetables
9	3803	23-12-2015	pot plants
10	2762	20-03-2015	whole milk
11	4119	12-02-2015	tropical fruit
12	1340	24-02-2015	citrus fruit
13	2193	14-04-2015	beef
14	1997	21-07-2015	frankfurter
15	4546	03-09-2015	chicken
16	4736	21-07-2015	butter
17	1959	30-03-2015	fruit/vegetable juice
18	1974	03-05-2015	packaged fruit/vegetables
19	2421	02-09-2015	chocolate
20	1513	03-08-2015	specialty bar
21	1905	07-07-2015	other vegetables
22	2810	08-09-2015	butter milk
23	2857	12-11-2015	whole milk
24	3962	18-09-2015	tropical fruit
25	1088	30-11-2015	tropical fruit
26	4976	17-07-2015	bottled water
27	4056	12-06-2015	yogurt
28	3611	13-02-2015	sausage
29	1420	14-01-2015	other vegetables
30	4286	08-03-2015	brown bread
31	4918	27-01-2015	yogurt
32	4783	22-10-2015	hamburger meat
33	3709	26-10-2015	root vegetables
34	4289	08-10-2015	pork

Groceries_dataset

Figure-4: Dataset

The figure-4 shows These are the Data that used to train a model.

V. RESULT

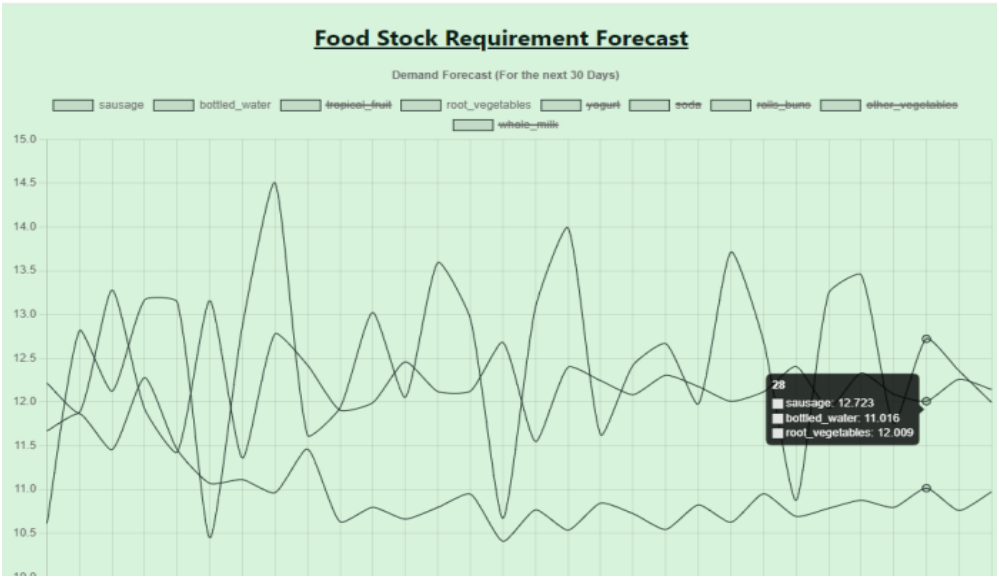


Figure-5: Items predicted for future days

The Figure-5 shows the Items predicted for future days, by plotting the forecasted demand based on days into the future, we can establish the nth-day where the demand becomes too low for any commercial value This value, n, will be used to determine the threshold.

V. CONCLUSION AND FUTURE WORK

In this paper, we are using external and internal data for the prediction consisting of different factors like region ID, week etc. Food demand prediction is an important and challenging problem. In this paper we presented penalized regression method, Bayesian Linear Regression K-nearest Neighbor, Decision tree approach as a food demand method. As we go through different algorithm for prediction the accuracy rate keeps on improving. There was not big difference other than precision rate of forecasting. XG-boost is a decision-based boosting algorithm which is used for increasing the accuracy rate.

This evaluation is used practically for restaurants. Furthermore, in future more refined prediction can be done based on many other factors like cultural habits, religious holiday, consumer preferences etc. In future, this method can be used for predicting work force requirement, automated food ordering based on forecasting results.

REFERENCES

[1] Tanizaki 2019 -"Demand forecasting based on a number of customers" .

[2] Wibowo and Yasmina 2021 - "Forecasts actual food consumption demand". Adv. Sci. Technol. Eng. Syst. J 6: 694–698.

[3] Bozkir and Sezer 2011 - "Forecasts actual food consumption demand". International Journal of Engineering Business Management.

[4] Tanizaki 2020 - "Forecasting customer order quantity of beers".

[5] Posch 2020 - "Illustrates the demand prediction system by predicting future sales of daily sold quantity for the restaurants and canteens". Sci. Technol. Eng. Syst. J 6: 694–698.

[6] Fattat 2018 - "Demand Forecasting for food products in a food manufacturing industry". 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1, pp. 1771–1776.

[7]Manjunatha HT and AjitDanti. "A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" -Recent Trends on Image Processing and Pattern Recognition, Springer Nature Singapore,Pte Ltd. 2019, **RTIP2R 2018, CCIS 1035, pp. 1–8, 2019, ISBN 978-981-13-9181-1DOI-https://link.springer.com/chapter/10.1007/978-981-13-9181-1_36**.

[8] Manjunatha HT and AjitDanti. "Detection and Classification of Potholes in Indian Roads using Wavelet

Based Energy Modules"IEEE- **978-1-5386-9319-3/19© 2019 ,SCOUPS Nature**

[9]Manjunatha HT, AjitDanti, Arunkumar KL, Rohith D" Indian Road Lanes Detection Based on Regression and clustering using Video processing Techniques", 3rd International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R,2020). Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. Springer, Scopus Indexed 3rd and 4th January 2020.**Springer Nature Singapore Pte Ltd. 2021 K. C. Santosh and B. Gawali (Eds.): RTIP2R 2020, CCIS 1380, pp. 1–14, 2021. https://doi.org/10.1007/978-981-16-0507-9_17**

[10] Manjunatha HT andAjitDanti," Indian traffic sign board recognition using Normalized Correlation Method", International Journal of Computer Engineering and Applications **(IJCEA), Volume XII, Issue III, March 18, ISSN 2321-3469**

[11] Manjunatha HT and AjitDanti, "Segmentation of Traffic Sign Board in a cluttered background using Using Digital Image Processing", **National Conference on Network Security, Image Processing and Information Technology, March 2017**.

[12] Estimation of vehicle distance based on feature points using monocular vision", Arunkumar K L, Ajit Danti, Manjunatha H T, IEEE 8816996 (2019), 1-5

[13] Peruse and Recognition of Old Kannada Stone Inscription Characters, CM Nrupatunga, KL Arunkumar, Springer, Singapore, 2020