IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.5, May 2022 E-ISSN: 2347-2693

Food Demand Forecasting Using Machine Learning And Statistical Analysis

Monika Agarwal[1*], Samarth Kulkarni[2], Vaishnavi Nagre[3], Aanchal Joshi[4], Damini Nagpure[5]

[1,2,3,4,5]Department of Computer Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Kopargaon, India

[*]*Corresponding Author: monikaagrawal@sanjivani.org.in, Mob. + 91-7709003254*

DOI: https://doi.org/10.26438/ijcse/v10i5.2529 | Available online at: www.ijcseonline.org

Received: 24/Apr/2022, Accepted: 10/May/2022, Published: 31/May/2022

Abstract— Food loss is considered a problem because food loss refers to the loss of resources such as water, soil nutrition, and investment. Food shortages lead to food shortages. This means that poor people around the world are being deprived of food as the cost of available food is increasing. Providing fresh food is one of the major constraints which is already considered by various meal provider agents or companies. Many of them want to get an estimated number of stocks for given respective times, which could help them understand patterns and stocks required.
Meal delivery companies want to know the estimated number of stocks that would be delivered or manufactured over the given period based upon previous data. Forecasting process is useful in various domains like weather forecasting, restaurants, retailing etc. It determines the expected demand for the future and establishes the level of readiness required on the supply side to meet the demand. This paper represents machine learning algorithms as an application to solve such problem with forecasting number of orders for given week and meal using algorithms Random Forest, XgBoost, Support Vector Machine, etc. with optimized results.

Keywords— Machine Learning, Prediction, Random Forest, XgBoost, Support Vector Machines, Clustering.

I. INTRODUCTION

The famous restaurant is known for its food and services, one of the vital features is keeping food fresh and managing its stock and storage. Also considering reach of customers, restaurant need to be prepared with all the raw materials needed for the same. It's very difficult to estimate number of orders for a given day or week for management of food items required for an order. Any mistake in that estimation would lead to either food wastage or food shortage, that's why predicting most accurate and approximate number of stocks is a real challenge based upon previous data.
There are many factors upon which this goal depends such as previous data, customer preferences, area for that food centre, food prices and many more.

That's why the number of orders keeps fluctuating day wise or week wise which makes this problem much more vital and essential to solve.
This research is mainly about finding and working on such problem for predicting the demand for food i.e., number of orders using previous given data from client up to 145 weeks.

This paper describes a solution for raised problem using machine learning. Proposed paper contains various section describing different aspect of it, such as Section II describes all related works are there, Section III contains Methodology used for research, Section IV indicates Architecture and Experiment Design, Section V states about Results and Observations where Section VI describes conclusion and Future Scope related to the same research.

II. RELATED WORK

Target is to predict number of orders for given data. As this problem falls under regression category which deals with a continuous value output, there are certain methods or algorithms which are built for same purpose. Random Forest Regressor, XGBoost Regressor, Support Vector Regressor are the three methodologies to be considered.

Decision Tree

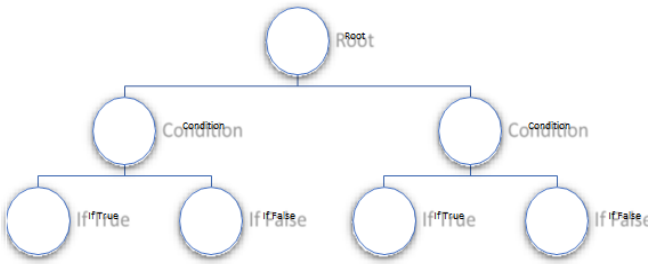

Fig.1

Based on certain condition it makes decisions. A decision tree is a kind of structure that looks like a hierarchical tree system. It is based on certain conditions, has a node as a function, and has parameters that are used to make some

decisions that lead to decisions. There are nodes and branches that represent different states of the decision.(Fig.1)

Regression Using Random Forest
It is possible to use decision tree in random forest for both classification and regression. In classification it is clear decision that which class it is but for regression it uses average method to decide final decision, Random Forest is a bagging technique in which multiple decision tress are used with either random based or grid-based selection of data and collectively on majority basis final decisions are made. This approach makes model stable and more accurate as multiple decision makers are used out of which most frequent decision is made.

XGBoost Regressor
The XGBoost algorithm is an implementation of a gradient decision tree. Engineered for performance and speed. Basically, XGBoosting is a kind of software library. So you have to access it through different interfaces. XGBoosting supports custom objective functions through classification, regression, and classification operations.
Silent features-

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

An objective function was used to measure the performance of the generated instance model. That is, a certain set of parameters is set. It also supports custom rating metrics.

- The XGBoost algorithm uses a gradient boosting decision tree algorithm.
- Gradient Boosting creates a new model that does the job of predicting the errors and residuals of all previous models, and then adds them together for final prediction.

Support Vector Regressor
The Support Vector Machine or SVM algorithm is a commonly used supervised learning algorithm for Classification, Regression, and outliers' detection problems, but in machine learning mostly SVM algorithm is used for classification. SVM algorithm classifies data into the classes by creating the hyperplane, and with help of this plane, new data will be put into the appropriate class. This hyperplane is called the best line or decision boundary. According to the distribution of the data, we can use Linear SVM and Non-Linear SVM these are the two types of SVM. As SVM works well for high dimensional spaces and our dataset consists of seasonal weekly data, therefore, we considered the SVM algorithm in our model. Implementation Steps of SVM algorithm: -

Step1: Provide the dataset as input.
Step2: Perform data pre-processing technique to make dataset reliable.
Step3: Try to visualize and analyse the data for making some decisions.
Step4: Divide dataset into Training and Test set.
Step5: Apply the SVM algorithm to the training dataset.
Step6: Test the algorithm
Step7: Check the Prediction accuracy result.

III. METHODOLOGY

Regular way
The regular way of training a machine learning model is application of an algorithm over whole training data, which leads to train a model considering residuals from every data that has passed to it. The general process consists of

1. Data Collection
2. Data Cleaning and Pre-Processing
3. Data Analysis and Visualization
4. Model Training (over whole data)
5. Result Analysis

Clustering Approach

- This approach consists of using clustering at the very first to group whole data into appropriate number of clusters
- Based of clusters all required algorithms are trained for each single cluster and out of those whichever algorithms suits that particular cluster is set as best algorithm for respective cluster
- K-means Clustering algorithms is used for clustering purpose and using *Elbow Method* it was decided that appropriate number of cluster that can be formed over given data was *3*.

Elbow Method
The most popular method for finding optimal number of clusters for provided data. It uses a range of values from 1 to K (where K is number of clusters) and calculates WCSS (Within Cluster Sum of Square) error for each K.
WCSS : It is a distance of a data point from a centroid of respective cluster

The elbow method requires drawing a line graph between the WCSS and the number of clusters (K) and finding a point representing the "elbow point". This is the point at which the WCSS, or inertia, begins to decrease linearly.(Fig.1)

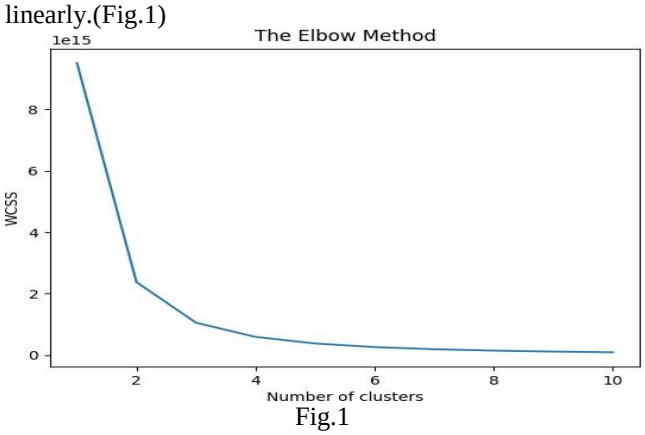

Fig.1

IV. ARCHITECTURE AND EXPERIMENT DESIGN

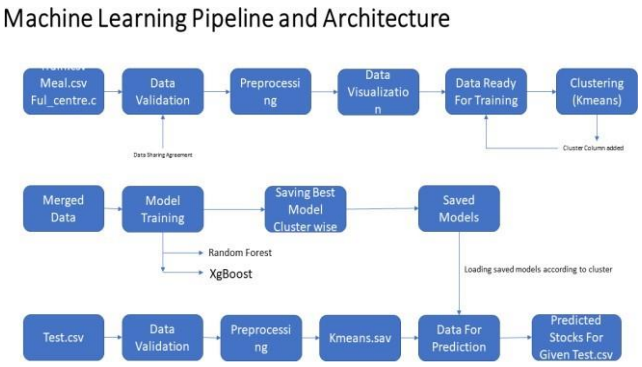

Fig.2

- Data Validation consist of validating given parameters (names, columns, rows) with provided Data Sharing Agreement.
- Pre-processing and Visualization generally involves detecting NULL values, checking correlations ,plotting graphs and getting some conclusions for further process.
- As Discussed about clustering approach, fisrt clustering has applied overl whole data and optimized number of clusters were found to be 3(using Elbow Method).
- ML modes were trained using algotithms and best suited algorithm was kept for that respective cluster (Fig.2)

V. RESULTS AND OBSERVATION

Before applying algorithms over data, we need to understand and pre-process data in detailed way. For this purpose, we have performed Data Visualization and observed following points-

- Type_A Centres have the highest number of Orders placed and Type_C has the least. (Fig.3)
- Type_A has the most number of orders because, Type_A has the most number of Centres - 43 Centres. (Fig.4)
- Italian Cuisine has the highest number of orders with Continental cuisine being the least. (Fig.5)

Respective Plots –

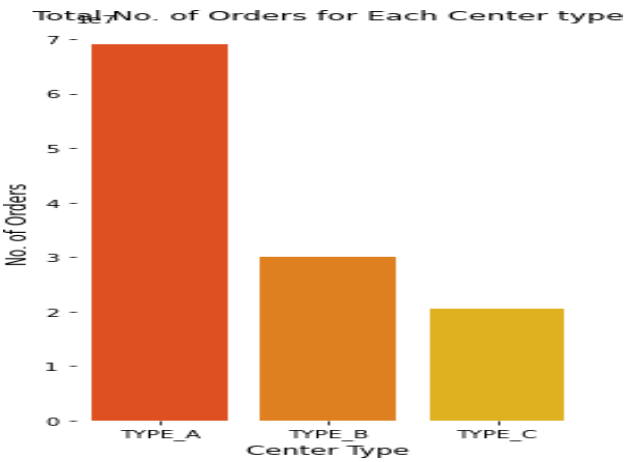

Fig.3

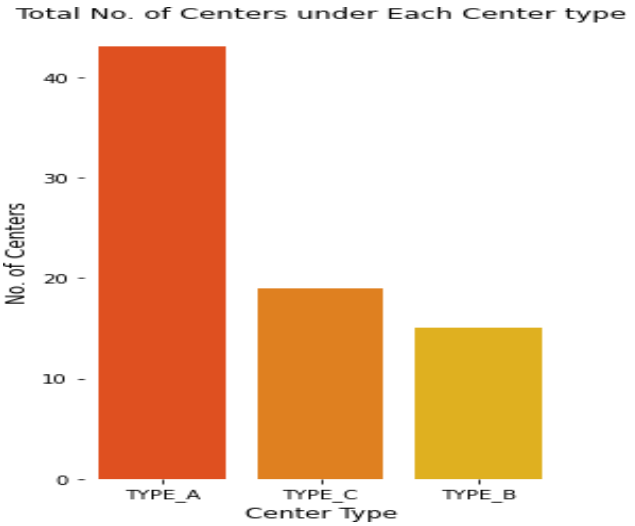

Fig.4

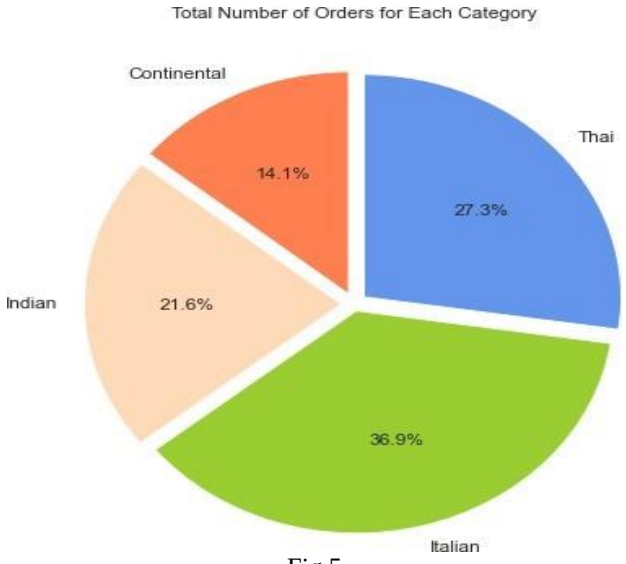

Fig.5

We created a new feature: Discount which is the difference of base price and checkout price and tried to find out if there is any relationship between the discount and the number of orders. But surprisingly there are no good correlation between the discount and the number of orders.

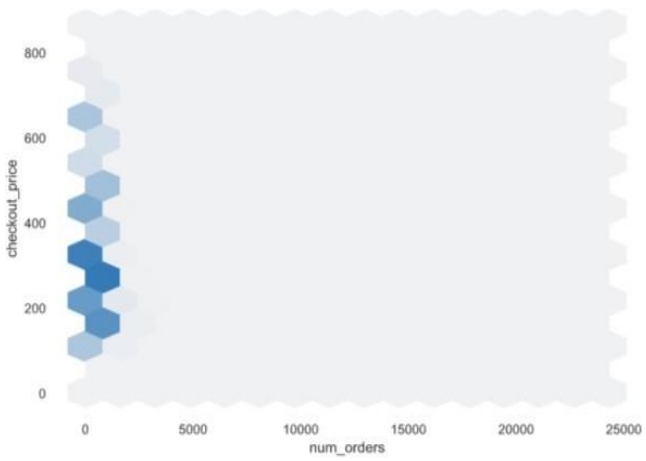

Fig.6

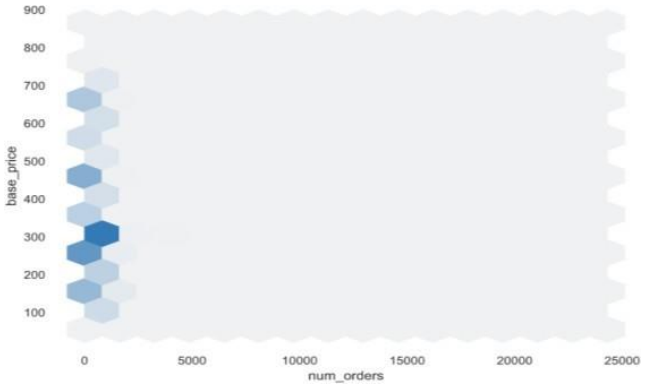

Fig.7

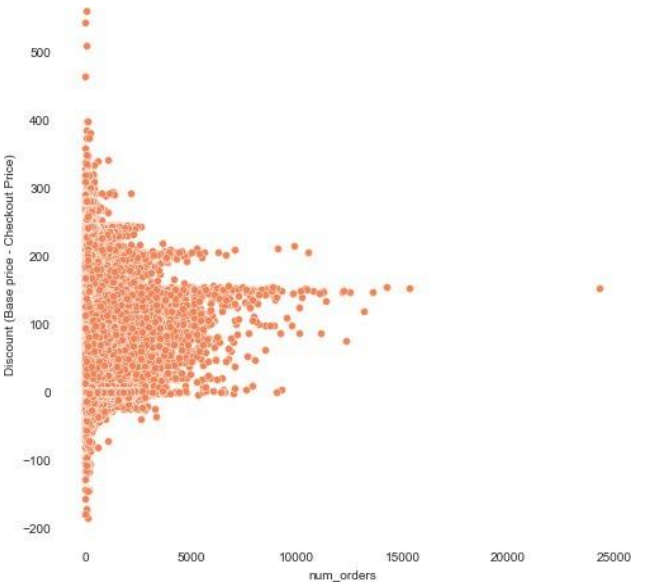

Fig.8

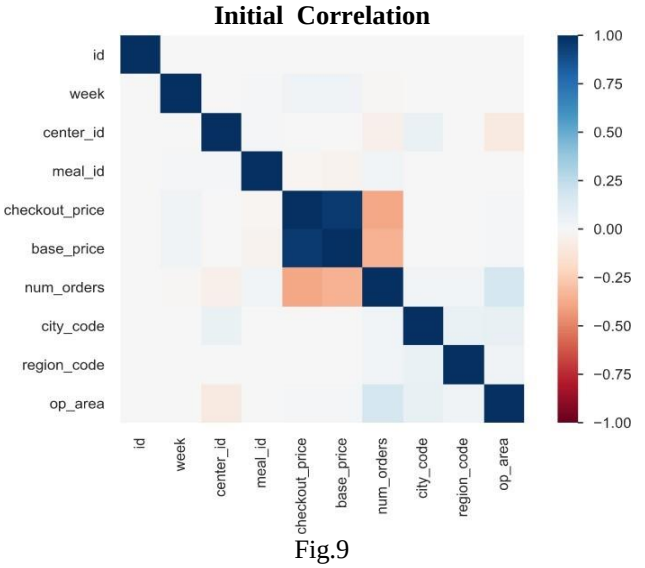

Fig.9

As we can see in Fig.9 there is very less correlation between checkout price and number of orders as well as between with base price (Fig.6 and Fig.7) . So derived a new feature called as discount and it shows clear relation with number of orders (Fig.8).
There are many features which have neutral relationship.

Results

As a part of results, the model was able to predict the number of orders for requested and given next 10 weeks of input data as a test validation data. All the three algorithms were able to make prediction model and out of them Random Forest Regressor was able to outstand as a robust and reliable predictive model for the same.
Results of statistical analysis were helpful in finding and deriving new features which had contributed in making reliable train data.

- Number of clusters from K-means – 3
- Cluster-1: Random Forest -78.3% (Highest)
  Cluster-2: Random Forest - 79.3%(Highest)
  Cluster-3: Random Forest - 81.3%(Highest)
- Out of all, Random Forest was the one performed well with better accuracy 81.3%.
- Apart from this, we have got some important conclusions which are discussed above.

VI. CONCLUSION AND FUTURE SCOPE

There is a huge demand and need of Data Science in Food Industry as meals are one of important basic need for a human being. Reducing food wastage is an challenge and absolute way of making profit for the food delivery or manufactures. Manufacturing of raw products generally depends upon features like price, discount, area and many things. Using machine learning it is possible to get estimated number of orders for a given week. One of the approaches is discussed in this paper where use of machine learning algorithms like Random Forest, XGBoost and SVM has worked in the favour and led to the decided goal.

For further future aspects many more terms need to be considered like customer reach, feedback, cultural habits, religious holiday, consumer preferences as well as use of Neural Networks would lead to more understanding and predicting accurate number of stocks and customer reach considering many more features . In future, this method can be used for predicting work force requirement, automated food ordering based on forecasting results. As a result, the system will push the company towards effective distribution of meals in the best possible way.

ACKNOWLEDGMENT

The food industry has using various AI applications for their betterment and business purpose, also helps in evolving with new discoveries leading to useful applications. We would like to extend our thanks to Prof.M.Agrawal,(Associate Prof., Computer Department, SCOE Kopargaon) for her valuable guidance and advice during the research. She was able to extend our learning capabilities towards the domain and provided us various opportunities to explore. This contribution helped us understanding industry level as well as subject knowledge which will be helpful to us during future events. Further, we are also thankful to Dr.D.B.Kshirsagar(H.O.D Computer Department) and Dr.A.B.Pawar (Project and Research Coordinator SCOE Kopargaon) for their valuable and considerable feedback.

REFERENCES

[1] D. Adebanjo and R. Mann, "Identifying problems in forecast-ing consumer demand within the fast paced commodity sector" Benchmarking: An International Journal, **7(3):223– 230, 2000.**

[2] Md.Erfanul Hoque, Aerambamoorthy Thavaneswaran, Srimantoorao S. Appadoo, "A Novel Dynamic Demand Forecasting Model for Rresilient Supply Chains using Machine Learning.",IEEE **ISSN:0730-3157,2021**

[3] K Siva Rama Krishna, Pooja Pasula, T.Kavyakeerthi, I.Karthik, "Identifying Demand Forecasting using Machine Learning for Business Intelligence.", IEEE, ISBN:**978-1-6654-1029-8, 2022**

[4] Kenji Shinoda, Masato Yamada, Motoki Takanashi, Tetsuya Tsuboi, "Prediction of Restaurant Sales during high demand states using population statistical data.", IEEE , ISBN:**978-1-6654-2397-7, 2021**

AUTHORS PROFILE

Prof. Monika Agarwal pursuing her PhD (CSE) from K L Deemed to be University, Vaddeswaram, A.P India, completed her ME (CTA) from SSCET Bhilai C.G and her BE(CSE) from BIT, Durg, C.G India. She is currently working as Assistant Professor in Computer Engineering Department, Sanjivani College of Engineering, Kopargaon and Maharashtra India. Her research interest includes Machine learning and Data Science. She has published/ presented more than 10 papers in various International Journals/Conferences also has 2 patents.

*Samarth M Kulkarni***,** pursuing his Bachelors Degree in Sanjivani College of Engineering, Kopargaon. His topics for research includes Machine Learning, Deep Learning, Data analytics, Predictive Algorithmns and Software Development.

*Aanchal S Joshi***,** pursuing her Bachelors Degree in Sanjivani College of Engineering, Kopargaon. Her research interest includes software development, data science and machine learning.

*Vaishnavi A Nagre***,** pursuing her Bachelors Degree in Sanjivani College of Engineering, Kopargaon. Research Topics are software development, cloiud computing and machine learning.

*Damini D Nagpure***,** pursuing her Bachelors Degree in Sanjivani College of Engineering, Kopargaon. Software development and machine learning are the topic she does research about.

