Food Demand Forecasting Using Advanced Machine Learning Modules

Snigdha Kandula[1], Ekta Lal[2]
[1, 2] Dept of Electronics and Communication Engineering
[1, 2] Vellore Institute of Technology, Chennai.

Abstract- *The vital aspect in the world of business is to have a proper analysis of their business outcomes. This outcome plays a major role in the development of the business. One of the expanding business spheres is food delivering companies. The vital factor in running such a food delivery company that is located at various branches in the city is to maintain the stock properly and prepare the food in time to deliver it to the customers. The Aim of the project is to develop a prediction model that predicts the Number of orders based on the unique id of the meal. The datasets used to train the model have the information regarding the meal for example, the type of meal, Week, center_id, base_price, category, cuisine etc. This prediction model helps to find out the popular meals and the least ordered type of meals, based on these results they can manage the purchase of stock and raw materials. They have to deal with many perishable raw materials, if there is too much stock it would tend to waste, if the stock is insufficient, it would lead to out-of- stocks, it would lead to out of stock of the meals that in turn would decrease the number of orders for the company. To develop the predictive model, we have used Linear Regression, Xgboost, Lightboost regressor, Catboost Regressor and Random Forest algorithm and tested this model on the test dataset to predict the Number of orders for upcoming weeks*.

Keywords- Data Analysis, Linear Regression, XG Boost, Light Boost Algorithm, Cat Boost Algorithm, Random Forest Algorithm.

I. INTRODUCTION

As in today’s competitive life, even the business has become more difficult. Demand for food is increasing day by day with the increase in the population of every country. Estimation of the demand in food consumption plays a vital role in supplying or generating resources to produce the required amount of food. To meet this challenge, we need to predict the demand in food consumption for the future so that the hunger of everyone can be satisfied. We will analyze all the previous year’s data on how the food demand has been throughout the restaurants. With that information, we can readily produce the raw materials and also recruit the staff required.

Food demand forecasting is about to help meal delivery companies located at various centers of the city in the demand forecasting for upcoming weeks. The majority of the raw materials are perishable, and the replenishment of them is done on a weekly basis and procurement planning is of utmost importance. The demand forecast is helpful in the staffing of the centers too. The Main Motto is to predict the demand for the upcoming weeks (in the challenge the test dataset contains 10 weeks) for the center-meal combinations in the test set. Here we will predict the number of orders and the demand of orders based on all the attributes given in the train data set.
If there is too much stock in the warehouse it tends to be a lot of wastage and if the stock is insufficient, it would lead to out-of-stocks and the number of orders received will get reduced and it pushes customers to seek solutions from other companies or your competitors. In this challenge, we will deal with a real dataset and the objective is to consider all the significant attributes in the data set and to train the model accordingly to predict the center-meal combination.

The motivation for this project was from the analytics Vidhya website, which gave us an amazing platform to work on our project and develop it. The resources and all the discussions on the platform have motivated and helped us in getting our project start and keep going.

II. RELEVANT WORKS

1. Barbosa et al.[1] have used holt winters method for time series analysis for effective forecasting.
2. Christo E et al.[2] focussed on finding the best forecasting model and then analysed the residuals control charts of the model.
3. Alexandrov T. et al.[3] have focussed on Model-Based Approach which assumes the specification of a stochastic time series model and showed how their properties can be improved by exploiting Reproducing Kernel Hilbert Space methodology.
4. Tanizaki, T. el al. [4] have used Bayesian Linear Regression, Boosted Decision Tree Regression, Decision Forest Regression and Stepwise method as the demand forecasting method.

5. Kuo, R. et al.[5] utilized a proposed fuzzy neural network (FNN), which is able to eliminate the unimportant weights and The result from FNN is further integrated with the time series data through an ANN.
6. Çetinkaya et al.[6] have proposed a model of artificial neural networks to estimate daily Meal demand
7. Anik, Asif et al. [7] have performed sensitivity analysis, based on past trends, we also hypothesized four alternative GDP and population growth scenarios and forecast corresponding changes in total energy demand forecast.
8. Rodrigues et al. [8] have researched about current machine learning algorithms for predicting food purchases considering temporal granularity of sales data, the input variables to employ for forecasting sales, and the representation of the sales output variable and also examines machine learning algorithms that have been used to anticipate food sales.
9. Gasparian et al.[9] have provided an analysis of the qualitative and quantitative methods of the demand forecasting and More complex techniques of time series include factors of trends, seasonal patterns, and economic cycles.
10. Kumar, Manoj et al. [10] have used Box-Jenkins' ARIMA model to forecast sugarcane production in India and found the order of the best ARIMA model and forecasted the future sugarcane production for a period upto five years .

III. PROPOSED SYSTEM

The implementation of the predictive modelling of Food Demand Forecasting undergoes several steps to give the accurate prediction. We used the RMSE (Root-Mean square error) as the evaluation metric of the model. Linear Regression, XGBoost, Light Boost regressor, Catboost Regressor and Random Forest Algorithms are used to build the model. The less the RMSE values, the more accurate is the prediction model. The steps involved in Proposed system are detailed as follows:

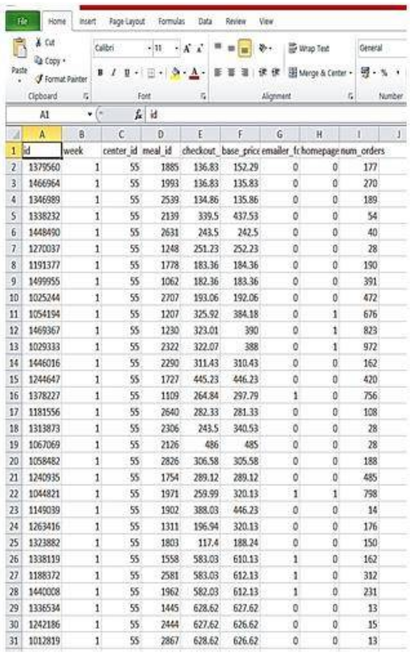
<u>Requirements</u>

The main requirements are the train data set, test data set and the sample submission which we can download from Kaggle or the analytics Vidhya portal. We use train and test datasets to train the models and test them with our predictions. Sample submission data set is used to check whether our predicted file is in the same format as the sample submission file. We have implemented this predictive model using Python programming using Jupyter Notebook.

<u>Description of Datasets</u>

There are a total of 3 train data sets which are used in the processing. The three data sets are the train data set, fulfillment centre info data set and the meal info data set. Train.csv data set includes all the attributes, the id's of the fulfillment centers and meals, and also the num_orders which is needed to be predicted, this is used to train the models, the fulfilment_center_info.csv file provides us with the details of the food centers which are providing the food. The meal_info.csv file contains all the extra and particular information about the means which are indicated with an id in the train.csv dataset. We have merged the train, fulfillment center and the meal info datasets to form one train dataset and used that to build the model. The attributes of each dataset are as follows:

Train Data set



	A	B	C	D	E	F	G	H	I	J
1	id	week	center_id	meal_id	checkout_	base_price	emailer_fc	homepage	num_orders	
2	1379560	1	55	1885	136.83	152.29	0	0	177	
3	1466964	1	55	1993	136.83	135.83	0	0	270	
4	1346989	1	55	2539	134.86	135.86	0	0	189	
5	1338232	1	55	2139	339.5	437.53	0	0	54	
6	1448490	1	55	2631	243.5	242.5	0	0	40	
7	1270037	1	55	1248	251.23	252.23	0	0	28	
8	1191377	1	55	1778	183.36	184.36	0	0	190	
9	1499955	1	55	1062	182.36	183.36	0	0	391	
10	1025244	1	55	2707	193.06	192.06	0	0	472	
11	1054194	1	55	1207	325.92	384.18	0	1	676	
12	1469367	1	55	1230	323.01	390	0	1	823	
13	1029333	1	55	2322	322.07	388	0	1	972	
14	1446016	1	55	2290	311.43	310.43	0	0	162	
15	1244647	1	55	1727	445.23	446.23	0	0	420	
16	1378227	1	55	1109	264.84	297.79	1	0	756	
17	1181556	1	55	2640	282.33	281.33	0	0	108	
18	1313873	1	55	2306	243.5	340.53	0	0	28	
19	1067069	1	55	2126	486	485	0	0	28	
20	1058482	1	55	2826	306.58	305.58	0	0	188	
21	1240935	1	55	1754	289.12	289.12	0	0	485	
22	1044821	1	55	1971	259.99	320.13	1	1	798	
23	1149039	1	55	1902	388.03	446.23	0	0	14	
24	1263416	1	55	1311	196.94	320.13	0	0	176	
25	1323882	1	55	1803	117.4	188.24	0	0	150	
26	1338119	1	55	1558	583.03	610.13	1	0	162	
27	1188372	1	55	2581	583.03	612.13	1	0	312	
28	1440008	1	55	1962	582.03	612.13	1	0	231	
29	1336534	1	55	1445	628.62	627.62	0	0	13	
30	1242186	1	55	2444	627.62	626.62	0	0	15	
31	1012819	1	55	2867	628.62	626.62	0	0	13	

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

The Fig 1 depicts the attributes of the train dataset in detail.

<u>Train Data set and it's variables</u>

Fulfillment centers Data set:

	A	B	C	D	E
1	center_id	city_code	region_co	center_typ	op_area
2	11	679	56	TYPE_A	3.7
3	13	590	56	TYPE_B	6.7
4	124	590	56	TYPE_C	4
5	66	648	34	TYPE_A	4.1
6	94	632	34	TYPE_C	3.6
7	64	553	77	TYPE_A	4.4
8	129	593	77	TYPE_A	3.9
9	139	693	34	TYPE_C	2.8
10	88	526	34	TYPE_A	4.1
11	143	562	77	TYPE_B	3.8
12	101	699	85	TYPE_C	2.8
13	86	699	85	TYPE_C	4
14	32	526	34	TYPE_A	3.8
15	149	478	77	TYPE_A	2.4
16	152	576	34	TYPE_B	4
17	92	526	34	TYPE_C	2.9
18	27	713	85	TYPE_A	4.5
19	14	654	56	TYPE_C	2.7
20	26	515	77	TYPE_C	3
21	104	647	56	TYPE_A	4.5
22	77	676	34	TYPE_A	3.8
23	23	698	23	TYPE_A	3.4
24	97	628	77	TYPE_A	4.6
25	146	526	34	TYPE_B	5
26	113	680	77	TYPE_C	4
27	145	620	77	TYPE_A	3.9
28	80	604	56	TYPE_C	5.1
29	55	647	56	TYPE_C	2
30	186	649	34	TYPE_A	3.4
31	99	596	71	TYPE_A	4.5

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)

The Fig 2 contains the variables and the dataset of the fulfilment centers data set taken from the analytics Vidhya portal. These are the details for the id's provided in the train data set undercentre_id.

Fulfilment data set and it's variables

Meal Info Data set

The Fig 3 contains the variables and the dataset of the meal info dataset taken from analytics Vidhya portal.These are the details for the id's Provided The train data set undermeal_id.

	A	B	C	D
1	meal_id	category	cuisine	
2	1885	Beverages	Thai	
3	1993	Beverages	Thai	
4	2539	Beverages	Thai	
5	1248	Beverages	Indian	
6	2631	Beverages	Indian	
7	1311	Extras	Thai	
8	1062	Beverages	Italian	
9	1778	Beverages	Italian	
10	1803	Extras	Thai	
11	1198	Extras	Thai	
12	2707	Beverages	Italian	
13	1847	Soup	Thai	
14	1438	Soup	Thai	
15	2494	Soup	Thai	
16	2760	Other Snac	Thai	
17	2490	Salad	Italian	
18	1109	Rice Bowl	Indian	
19	2290	Rice Bowl	Indian	
20	1525	Other Snac	Thai	
21	2704	Other Snac	Thai	
22	1878	Starters	Thai	
23	2640	Starters	Thai	
24	2577	Starters	Thai	
25	1754	Sandwich	Italian	
26	1971	Sandwich	Italian	
27	2306	Pasta	Italian	
28	2139	Beverages	Indian	
29	2826	Sandwich	Italian	
30	2664	Salad	Italian	
31	2569	Salad	Italian	

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

Fig 3

Meal Info Data set and it's variables

Preprocessing of the data

In this session, we have imported the datasets. We have come to the dimensions and the complete information about the datasets. We have checked the presence of duplicate values. We have assigned a random constant value to the target variable i.e.number of orders as we have to predict that column after training the model. we have merged the meal info, fulfillment Centre,test datasets with train dataset to form one train dataset. We used the data.isnull().sum() to find out the missing values and found that there are no missing values in the dataset.

Feature Engineering

In this section, we extract some useful features from existing attributes that helps in improving the performance of prediction model. We create Some Features that are as follows, discount amount, 'discount percent', compare_week_price, compare_week_pricey/n.

Exploratory Data Analysis

The section includes the data visualization, after getting an complete understanding on the dimensions and the attributes of the datasets.Through EDA we can analyze every single variable and their nature of distribution. We did the Bivariate Analysis on which we plotted everysing attribute with target variable. so we can find the relationship between independent and the target variable and can draw some useful inference.

Encoding Categorical Variables

The encoding of the categorical variables is essential as the machine learning models gives better prediction result with continuous and numerical variables. The complete removal of categorical variables would lead to loss of information. Here we used get_dummies() method to convert each category of a variable to a number.

Predictive Modeling

The most crucial stage and the heart of the project is the predictive modeling using machine learning algorithms. We used the Linear Regression Model, Xgboost, LightBoost regressor, Catboost Regressor and Random forest Algorithms implementing the project. We have scikit learn package and did the test train split to train the model and tested the montest dataset.The algorithms and the models build using them are briefed in Module description session.

Evaluation metric is the vital part of building a effective model as we get feedback from this metrics to improve the model further. Here, we used the RMSE (Root-Mean Squared Error) as evaluation metric.

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Root Mean Squared Error(RMSE):is the square root of the mean of the squared errors.

System Working

The implementation of the predictive modeling of Food Demand Forecasting undergoes several steps to give the accurate prediction. We used the RMSE (Root-Mean square error) as the evaluation metric of the model. Linear Regression, Xgboost, LightBoost regressor, Catboost Regressor and Random Forest Algorithms are used to build the model. The less the RMSE values, the more accurate is the prediction model.

IV. RESULTS

Data Visualization

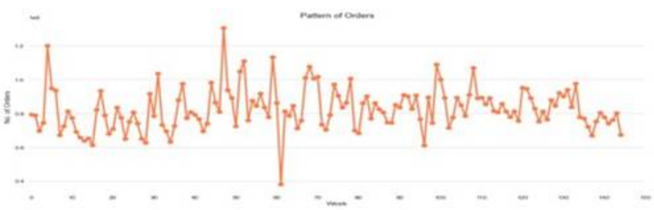


Fig 4 no_of_orders vs week plot

The Above Fig 4 shows that week 62 received low number of orders and week 5 and 48 received the highest number of orders.

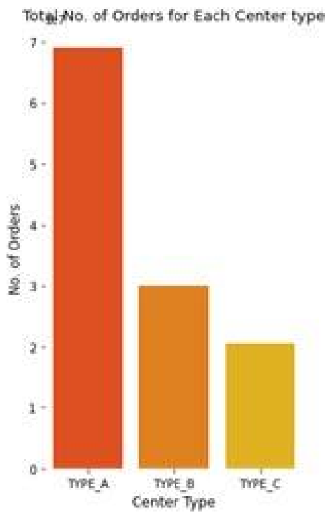


Fig 5 no_of_orders vs center_type plot

The above Fig 5 shows that Centre of Type_A has highest orders and Type_c have least orders.

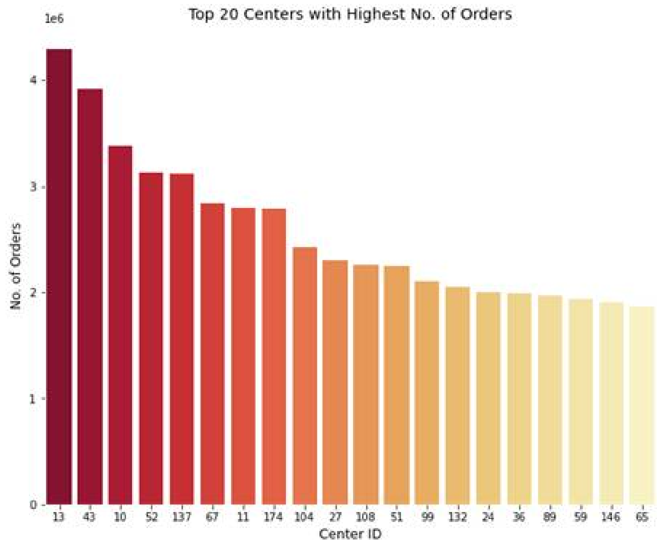


Fig 6 no_of_orders vs center_id

Previously when we analysed the target variable vs centre_type, type_a centre type receives a greater number of orders but in Fig 6 we can see that centre_id 13 that belongs to center_type B receives a greater number of orders

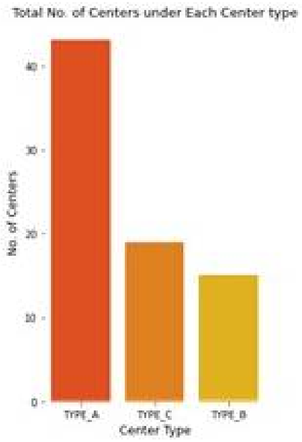


Fig 7 center_type vs no_of_centers

In the above Fig 7 it is observed that there are a greater number of centres of the type_As it resulted in greater number of orders in this type,but analysed individually the highest orders placed in other centre_id of type_B can be concluded.

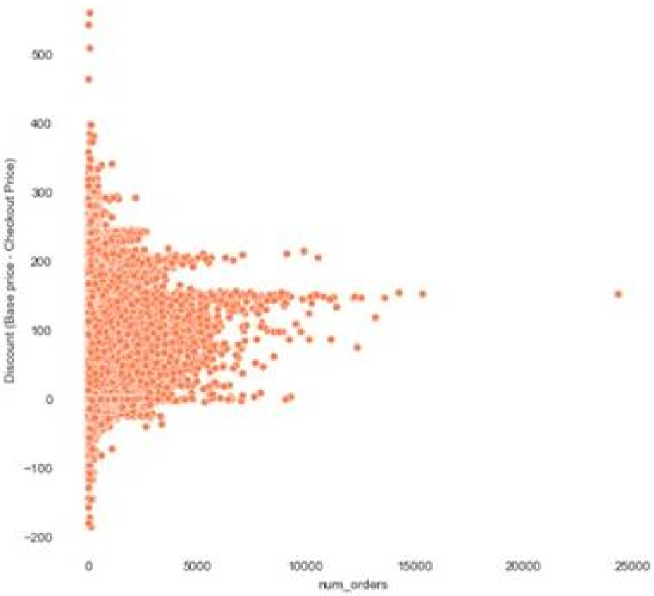


Fig 8 num_orders vs discount

Here in the above Fig 8, we can observe there is no good relation between number of orders and discount.

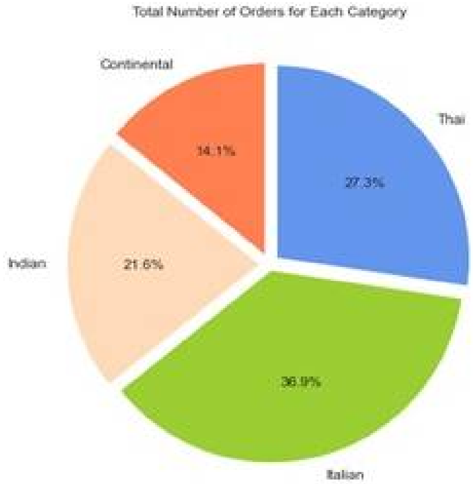


Fig 9 Pie Chart with no of order for each category

In the above Fig 9 the variation of cuisines can be observed as the total number of orders in each category.

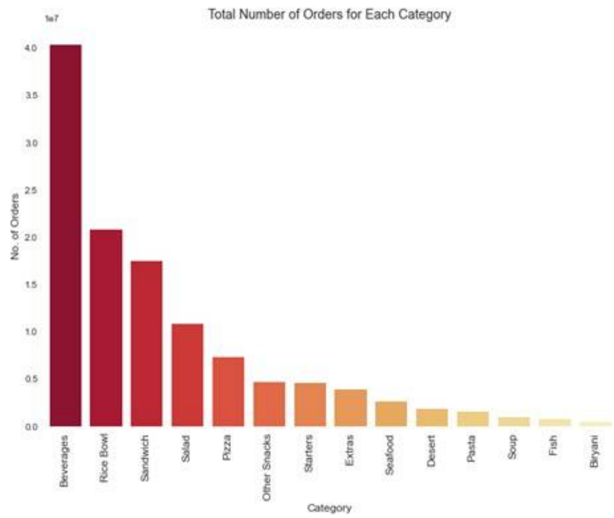


Fig 10 Category of foods vs no of orders

In the above Fig 10 we can observe that beverages have the high number of orders and the biryani have the least number of orders

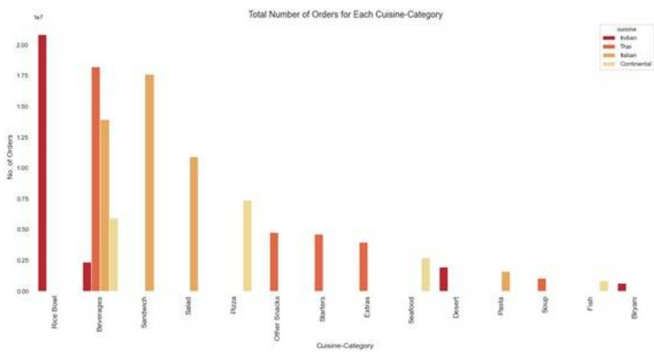


Fig 11 Subplot of cuisine category vs count

From the above Fig 11 we can observe that the rice bowl of Indian cuisine has the highest count of orders history and the biryani has the least count.

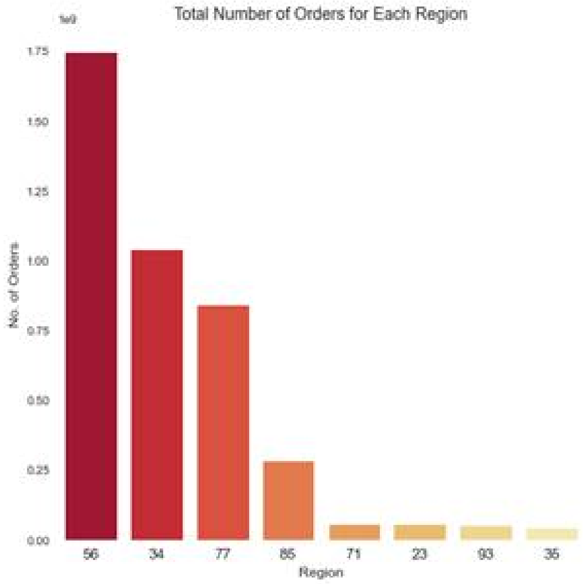


Fig 12 No of orders per region

From the above Fig 12 it is observed that the region with code 56 has the highest number of orders and the region with code 35 has the lowest number of orders.

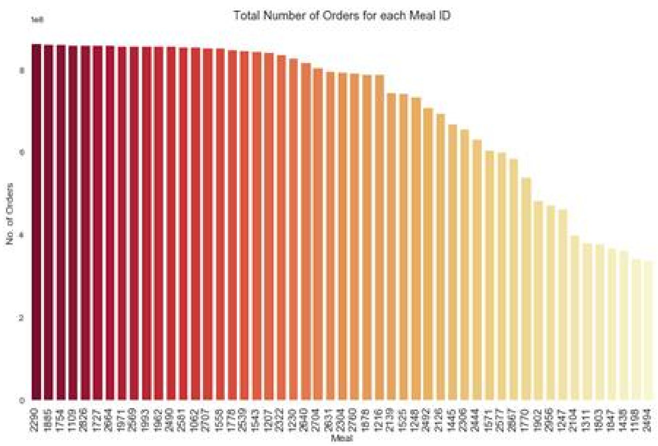


Fig 13 meal_id vs no_of_order

If we observe the Fig 13 there is not much observable difference in the meal_id and the number of orders. meal_id with 2290 received the highest number of orders.

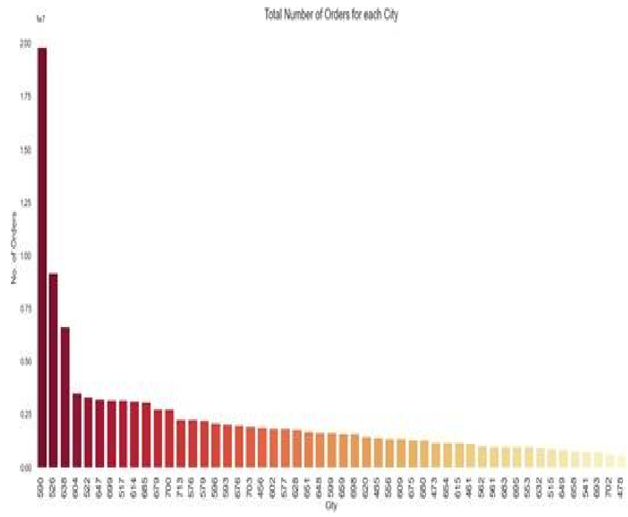


Fig 14 city vs no_of_orders in that city

On analysing the Fig 14 the number of orders we can observe that the city id with 590 has a greater number of orders and that is 2 times of order compared to second highest.

Comparison of RMSE values of Different Models is shown in Table 1

Table 1

S.NO	Algorithm	Model	RMSE value
1	Linear regression	Model 1	194.38
2	Linear regression	Model 2	0.6347
3	Linear regression	Model 3	0.6246
4	XGBoost	Model 1	0.5139
5	XGBoost	Model 2	0.5058
6	Light Boost Regressor	Model 1	0.5258
7	Light Boost Regressor	Model 2	0.5079
8	Light Boost Regressor	Model 3	0.5076
9	Light Boost Regressor	Model 4	0.5123
10	Random Forest Algorithm	Model 1	0.5289
11	Catboost Regressor	Model 1	0.4966
12	Catboost Regressor	Model 2	0.4960

RMSE score obtained on testing the model with test dataset is tabulated in the Table 2. We have submitted the obtained csv prediction files and the table below depicts the scores.

Table 2

S.NO	Algorithm	Model	RMSE value*100
1	XGBoost	Model 1	54.640
2	XGBoost	Model 2	54.0281
3	Light Boost Regressor	Model 1	54.0277
4	Light Boost Regressor	Model 2	54.4690
5	Light Boost Regressor	Model 3	53.7960
6	Light Boost Regressor	Model 4	53.9865
8	Catboost Regressor	Model 1	53.0917
9	Catboost Regressor	Model 2	52.5823

V. CONCLUSION

In the Present era of the growing business world, every company has to focus on each and every input and result obtained. Making an Analysis of their business can give a good profit and result. Using this Analysis, they can make changes in the way of development and can satisfy their customers. Broad examination around there at the big business level is occurring for precise deals forecast. As the profit made by an organization is straight forwardly relative to the precise expectations of deals, the companies want more exact forecast calculation with the goal that the organization won't su□er any misfortunes.

In our project, the company is Food Delivery Service Company. To run their business without any loss one of the vital aspects is to maintain the stock properly. Here, we developed a predictive model that predicts the number of orders using Linear Regression, XGBoost, Catboost Regressor, Light Boost Regressor, and the Random Forest Algorithms. In comparison of results of this model Catboost regressor model-2 has given the best prediction results with the least RMSE value.

Through this prediction model, they can get a better idea regarding the number of orders received in the upcoming weeks and they can plan the stock accordingly. If the number of orders is more for a particular, they can purchase those rawer materials needed for that meal and if the number of orders is less, they can decrease the raw material related to those meals. In this way, our prediction model can help the food delivery company to plan their stock accordingly and satisfy their clients.

REFERENCES

[1] Barbosa, Nathalia & Christo, Eliane & Alonso Costa, Kelly. (2015). Demand forecasting for production planning in a food company. 10. 7137-7141.

[2] Christo E., Ferreira M. and Alonso K. 2013. 'Use of Statistical Control for Improved Demand Forecasting', Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC), 2013 BRICS Congress on. IEEE.

[3] Alexandrov T., Bianconcini S., Dagum E. B., Maass P. and Mcelroy T. S. 2012 'A review of some modern approaches to the problem of trend extraction', Econometric Reviews, Vol. 31, pp. 593-624.

[4] Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. *Procedia CIRP*, *79*, 679–683. https://doi.org/10.1016/j.procir.2019.02.042

[5] Kuo, R., Wu, P., & Wang, C. (2002). An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination. *Neural Networks*, *15*(7), 909–925. https://doi.org/10.1016/s0893- 6080(02)00064-3

[6] Çetinkaya, Zeynep & Erdal, Erdal. (2019). Daily Food Demand Forecast with Artificial Neural Networks: Kirikkale University Case. 1-6. 10.1109/UBMK.2019.8907105.

[7] Anik, Asif & Rahman, Sanzidur. (2021). Commercial Energy Demand Forecasting in Bangladesh. Energies. 14. 10.3390/en14196394.

[8] Silva, Juliana & Figueiredo, Manuel & Braga, Ana Cristina. (2019). Demand Forecasting: A Case Study in the Food Industry. 10.1007/978-3-030-24302-9_5.

[9] Rodrigues, Aaron. (2021). Food Sales Forecasting Using Machine Learning Techniques: A Survey. International Journal for Research in Applied Science and Engineering Technology. 9. 869-872. 10.22214/ijraset.2021.38069.

[10] Gasparian, Mikhail & Karmanov, M.V. & Kiseleva, I.A. & Kuznetsov, V.I. & Sadovnikova, N.A.. (2018). Modeling of the demand forecasting. International Journal of Civil Engineering and Technology. 9. 163-173.

[11] Kumar, Manoj & Anand, Madhu. (2014). An Application Of Time Series Arima Forecasting Model For Predicting Sugarcane Production In India. Studies in Business and Economics. 9. 81 - 94.

[12] Junior M. L. and Filho M. G. 2012. 'Production planning and control for remanufacturing: literature review and analysis', Production Planning & Control, Vol. 23, pp. 419-35.

[13] Box, G.E.P. & Jenkins, G.M.. (1971). Time Series Analysis, Forecasting and Control. Journal of the American Statistical Association. 134.

[14] Tratar, L. F. (2015) 'Forecasting method for noisy demand', International Journal of Production Economics, 161, 64-73.

[15] Taylor, David & Fearne, Andrew. (2009). Demand management in fresh food value chains: A framework for analysis and improvement. Supply Chain Management: An International Journal. 14. 379-392. 10.1108/13598540910980297.

[16] Lewis C. 1997. “Demand forecasting and inventory control”, Control-Coventry-Institute Of Operations Management, Vol. 23, pp. 20-23.

[17] Cecatto, Cristiano & Belfiore, Patrícia. (2015). Demand forecasting methods in the Brazilian food industries. Gestao e Producao. 22. 404-418. 10.1590/0104-530X108-12.

[18] Wood, Frances & Lee, Susan. (2021). Modelling food demand in the 21st century. Food Science and Technology. 34. 43-46. 10.1002/fsat.3403_11.x.

[19] Valin, Hugo & Sands, Ronald & Van der Mensbrugghe, Dominique & Nelson, Gerald & Ahammad, Helal & Blanc, Elodie & Bodirsky, Benjamin & Fujimori, Shinichiro & Hasegawa, Tomoko & Havlík, Petr & Heyhoe, Edwina & Kyle, Page & Mason-D'Croz, Daniel & Paltsev, Sergey & Rolinski, Susanne & Tabeau, Andrzej & Meijl, Hans & von Lampe, Martin & Willenbockel, Dirk. (2014). The Future of Food Demand: Understanding Differences in Global Economic Models. Agricultural Economics. 45. 10.1111/agec.12089.

[20] The Combination of Forecasts 1969J. M. Bates, C. W. J. Granger10.1057/jors.1969.103 Journal of the Operational Research Society