FINAL DELIVERABLES

Team id	PNT2022TMID32327
Project Name	DemandEst – AI Powered Food Demand Forecasting

INDEX

ABSTRACT

The food and beverage industry is one of the most important sectors of the nation economy, with a significant participation in GDP index. The economy has been showing a relative stability in the last decades, which takes the sales demand to be more predictable. Due to this scenario of economic stability, the companies has been worried about investing in planning their operations, making use, mainly, of forecasting methods in order to become more competitive in the market. In the case of food industry, the seasonal and the short perishability factors are a limitation to the maintenance of stocks, requiring a forecast with a high accuracy level. The present work consists in applying methods to forecast the demand for products of a food industry, which directs its sales to the food service market, in order to base the short to medium term production planning. Posteriorly, the forecasts will be evaluated using the error measure MAPE and compared to the demand currently considered by the company. The proposed methods feature a reduction of the error approximately 5%.

Keyword: *Prediction, economy, demand, forecast*

1.INTRODUCTION:

A food delivery service has to deal with a lot of perishable raw materials which makes it all, the most important factor for such a company is to accurately forecast daily and weekly demand. Too much inventory in the warehouse means more risk of wastage, and not enough could lead to out-of-stocks - and push customers to seek solutions from your competitors. The replenishment of majority of raw materials is done on weekly basis and since the raw material is perishable, the procurement planning is of utmost importance, the task is to predict the demand for the next 10 weeks.

The main aim of this project is to create an appropriate machine learning model to forecast the number of orders to gather raw materials for next ten weeks. To achieve this, we should know the information about of fulfilment center like area, city etc., and meal information like category of food sub category of food price of the food or discount in particular week. By using this data, we can use any classification algorithm to forecast the quantity for 10 weeks. A web application is built which is integrated with the model built.

These variations ad fluctuations in demand may be because of price change, promotions, change in 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 regular seasonal pattern is expected, the features that predict these seasons are not directly observed. Thus, drops and rises in orders because of these seasonal changes are difficult to predict. In order to solve such problems, we are researching how to predict the demand [1]. Here we are researching food demand forecasting methods using internal data such as number of orders.

The demand forecasting methods can be based in mathematical models that use historical data or in qualitative methods, planned according to the administrative experience and customers reviews. They can also be based in a combination of both quantitative and qualitative methods. The demand forecasting methods can be based in mathematical models that use historical data or in qualitative methods, planned according to the administrative experience and customers reviews. They can also be based in a combination of both quantitative and qualitative methods.

2. FORECASTING MODELS:

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.

2.1 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.

2.2 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.

2.3 The Bayesian network

It 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 that variable.

2.4 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 .

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

2.6 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. The lasso procedure encourages simple, sparse models . This particular kind of regression is well-suited once we would like to automate certain parts of model selection, like variable selection/parameter elimination.

2.7 XGBoost

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 with very less computation time, leading to better performances with real data . The amount of incorrect predictions is increased to enhance the accuracy of the training model. XGBoost is an implementation of gradient boosted decision trees which is designed for speed and performance. XGBoost 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 treesDF.
- A proportional shrinking of leaf nodes.
- Newton Boosting.
- Extra randomization parameter

3. FORECASTING OF NUMBER OF CUSTOMERS

3.1 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 XGBoost to solve the problem. Hence, if we observe weekly and 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.

3.2 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. 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. XGBoost algorithm has given the highest accuracy and thus gives us better performance with respect to other models.

4. IN FOOD INDUSTRY

4.1 Impacts of Demand Forecasting Errors

For companies, mainly large ones with economies of scale and geographic capillarity, an error in the forecast of demand can cause several consequences, such as:

- Stock break;
- Perishable waste (What is FIFO?);
- Drop in production;
- Idle stock (slow moving)
- Pricing errors

Adversities like these directly impact the companies' final results, as they result in loss of market share, increase in costs or low optimization in the dilution of fixed costs, growth in the loss of perishable products, frustration of employees in relation to the goals and mainly break in the confidence of recurring customers who depend on supply for their operations.

4.2 The demand forecast in the food sector

The food industry is situated in a context of highly perishable products with the following characteristics:

- High inventory turnover;
- Parallel supply in different locations;
- Large number of Skus, points of production and points of sale;
- Verticalized supply chain;
- Non-linearity in data patterns;
- Seasonality.

These characteristics make the sector a business niche that is more sensitive to deviations in demand forecast and adjacent planning.

4.3 The role of machine learning in the food sector

The use of AI through machine learning techniques associated with a coherent technological stack of analytics .Provides greater information speed, data organization with different granularities (region, state, city and neighborhood), adjustments seasonality, exploration of opportunities and decision making in real time.

In the case of the food sector, the greatest accuracy in forecasting demand means:

- Inventory optimization among Distribution Centers (CDs);
- Reduction of idle stocks;
- Decrease in disruptions that cause loss of market share due to substitute products;
- Direct reduction in losses with perishability (FIFO).

The great technical and conceptual challenge faced by data scientists (The profile of data scientists in the view of Aquarela), however, is the modeling of analysis datasets (what are datasets?) That will serve for the proper training of machines.

5.BUILDING APPLICATION FOR FOOD DEMAND FORECASTING

The food demand forecasting application is constructed using the basic web development languages. Here HTML is used to create the overview for an website and CSS is used to apply the style to the website for an attractive look.

HTML:

The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.

CSS:

Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

These files are being stored in the Flask File for the easy use of the codings.

This website contains the two main webpages which directs the users to find the demand for the particular warehouse.

5.1 HOME.HTML:

It contain the overview of the food demand forecaster website and the main use of the website.

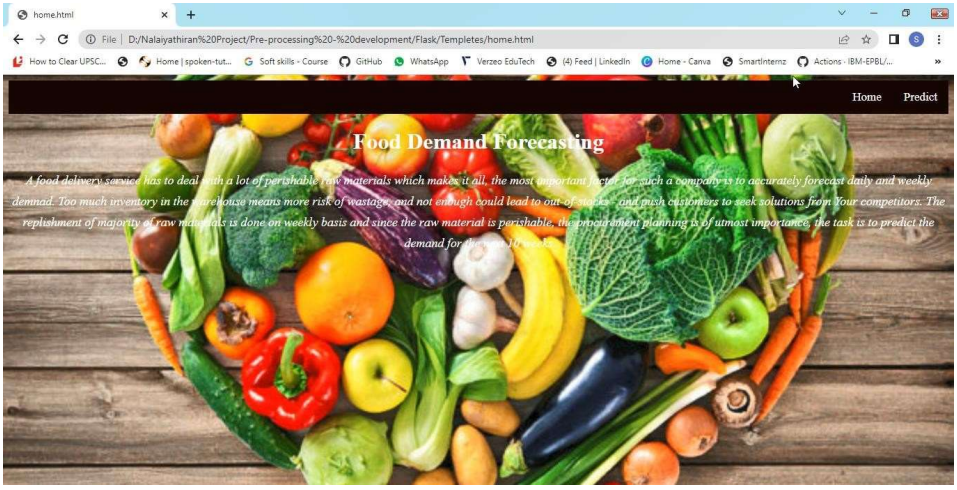

Figure 3.1: Home page

5.2 UPLOAD.HTML:

This webpage contain the information which should be got from the user for access the information like location, type of food, pin code, city , etc..

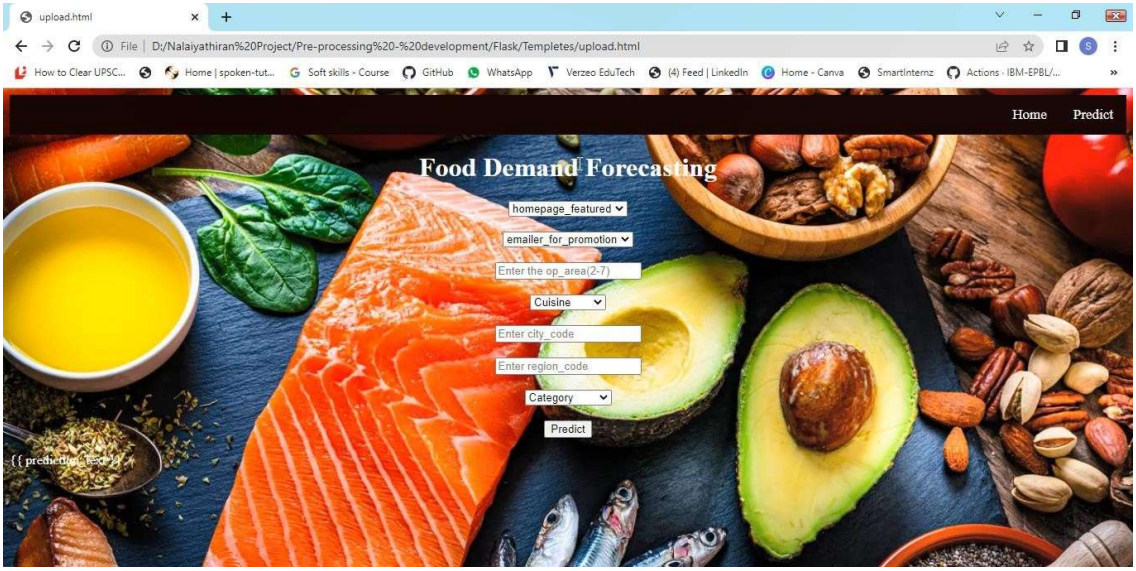

Figure 5.2: Upload page

Then the items predict will be retrieved by using the python program which is deployed by using the IBM Watson.

6.TRAINING BY IBM CLOUD

In IBM Watson Knowledge Studio , the creation of the machine learning model involves training the machine learning model and evaluating how well the model performed when annotating test data and blind data.

6.1 Creating a machine learning model

When you create a machine learning model, you select the document sets that you want to use to train the model and specify the percentage of documents that are to be used as training data, test data, and blind data.

Procedure

To create a machine learning model:

1. Log in as a Knowledge Studio administrator and select your workspace.
2. Select **Machine Learning Model** > **Performance**.
3. Verify that all of the document sets have been approved and that all annotation conflicts have been resolved through adjudication. Only documents that have become ground truth through adjudication or approval can be used to train the model.
4. Click **Train and evaluate**.
5. Optional: To specify how you want to allocate documents from your document sets to be used by the system-level training, test, or blind sets, click **Edit settings**.
6. Click **Train** to train the model, or click **Train & Evaluate** to train the model, evaluate annotations added by the machine learning model, and analyze the performance statistics.
7. Select the document sets that you want to use for training the model.
8. After the model is created, select one of the following actions:

Option	Description
Log	View the log file to see whether any problems occurred.
Details	View the annotation performance statistics, change the document sets that you want to use for training and testing the model, and create snapshot versions of the model artifacts.
Export	If you have a Standard plan or a Premium plan, you can export a file to your local system that contains the components that are required for the model to run in a machine learning runtime environment, such as Watson Explorer.

6.2 Evaluating annotations added by the model

You can compare the ground truth view for annotations added by human annotators to the annotations added by the model.

Procedure

To evaluate the annotations added by the model:

1. Select **Machine Learning Model** > **Performance** > **Train and evaluate**. The Training/Test/Blind Sets page is displayed.
2. Click **View Ground Truth** for the training set or test set to see the annotations that were added through pre-annotation and by human annotators. The ground truth editor opens. Click to open individual documents and see how the mentions, relations, and coreferenced mentions were annotated.
3. On the **Performance** page, click **View Decoding Results** to see the annotations that the machine learning model added to documents in the test set. This button is available only after you evaluate the model. By viewing results, you can see how well the machine learning model labeled mentions, relations, and coreferenced mentions in the test data.
4. If you want to change how the documents are divided between training, test, and blind data sets, click **Performance** > **Train and evaluate** > **Edit Settings**. For example, if initial results seem acceptable, you might want to increase the number of documents in the test set to further test the machine learning model's results. You can change the ratio for how documents are automatically divided for different purposes, or you can select specific document sets to use as training data, test data, and blind data.
5. If you made any changes, click **Train & Evaluate** to retrain the model and re-evaluate the annotations.

6.3 ANACONDA:

Anaconda Distribution equips individuals to easily search and install thousands of Python/R packages and access a vast library of community content and support. Anaconda is the birthplace of Python data science. We are a movement of data scientists, data-driven

enterprises, and open source communities. A powerful and versatile machine learning library for machine learning basics like classification, regression, and clustering. Anaconda is a distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment.

7. 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. XGboost 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 forecastingresult

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 forecast-ing 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] İrem İşlek and Şule Gündüz Öğüdücü," A Retail Demand Forecasting Model Based on Data Mining Techniques".

[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-reinedd9f99be63d.