1. Introduction

1.1 Overview

There is a meal delivery company that operates in multiple cities. They have various fulfillment centers in these cities for which dispatch meal orders to their customers. The project aims to help these centers with demand forecasting for upcoming weeks so that these centers can plan the stock of raw materials needed accordingly.

The replenishment of the majority of raw materials is done on weekly basis and since the raw material is perishable, procurement planning is crucial. Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful. Our Machine Learning model will predict the demand for the upcoming 10 weeks for the center-meal combinations in the test set.

1.2 Purpose

To create a Machine Learning algorithm that predicts the quantity of raw materials required in a given Region and City for a week.

To optimize our resources and avoiding wastage using digital technology to help you accurately predict the number of food orders

To obtain a better understanding of food order patterns, and more control over the enterprise, which ultimately leads to less wastage and thus more savings.

2. Literature Survey

2.1 Existing Problem

Mismanagement/wastage of food leads to increased food inequity for the poor and also damage to the environment. Existing solutions that are implemented include:

- Measuring food waste and based on that implementing the necessary changes

- Engaging the staff for getting innovative solutions to tackle the problem
- Practicing stock control, clearly labeling, and organizing stock according to expiry dates
- Avoiding batch preparation and overproduction
- Composting
- Recycling
- Donating to the homeless and others in need

2.2 Proposed Solution

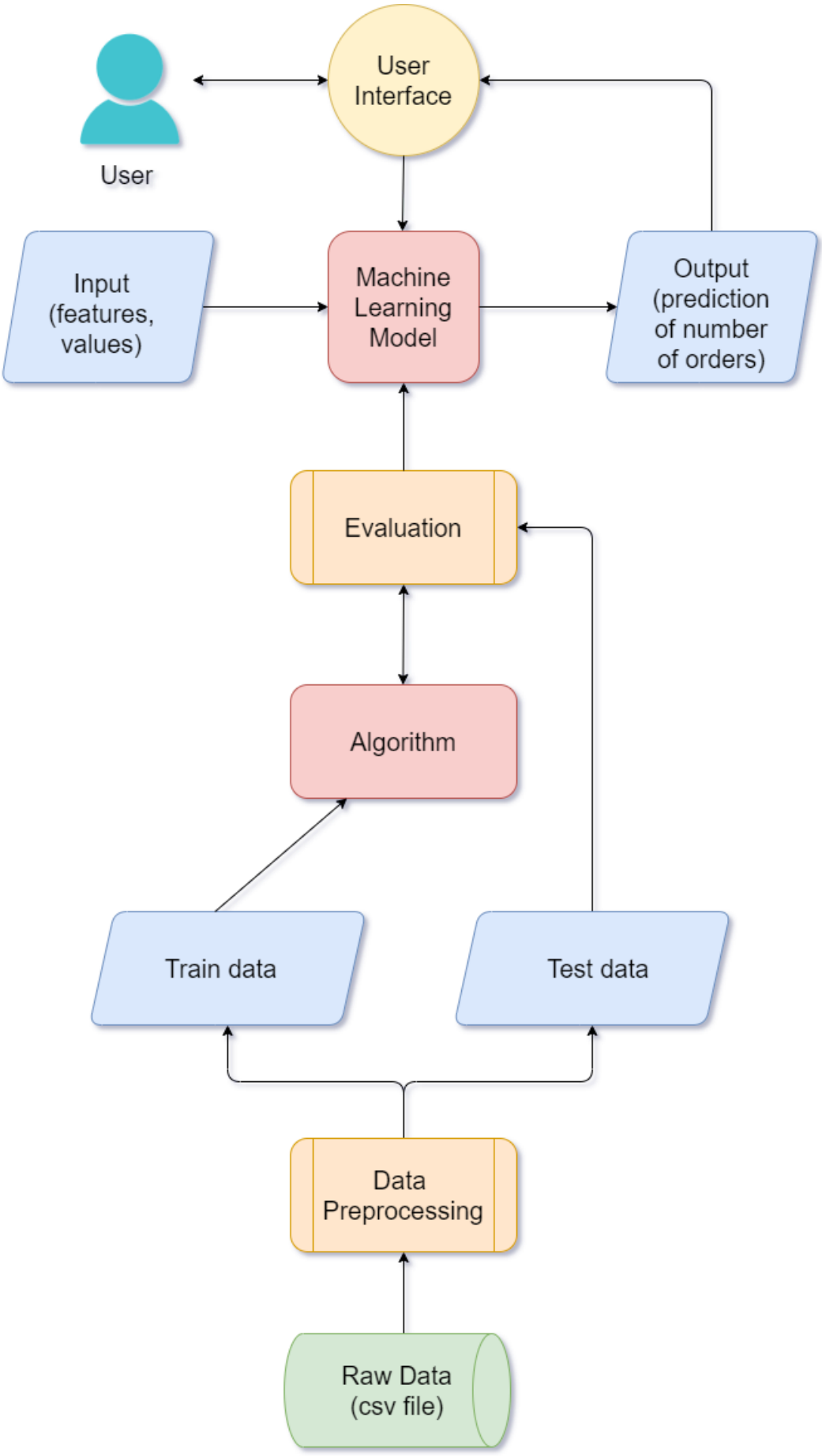
Having a system in place to help you accurately predict food orders with the help of digital technology means more accurate data, a better understanding of food order patterns, and more control over the enterprise, which ultimately leads to less wastage and more savings.

3. Theoretical Analysis

3.1 Block diagram

This block diagram on the following page gives us a diagrammatic overview of the project.It explains how different components and processes interact with each other within this project to give us the final outcome.

Raw data is transformed in data processing after which it gets split into training and testing data.
The training data is used to train the machine learning algorithm. The algorithm is used in the process of evaluating the testing data.
The testing data has the same features as the training data except for the target variable (number of orders) which is to be predicted. The testing data is sent for evaluation process.
The Machine Learning model takes in the input from user i.e features, values and gives an output i.e the number of orders.
The output gets displayed to the user on the user interface.


User
User Interface
Input (features, values)
Machine Learning Model
Output (prediction of number of orders)
Evaluation
Algorithm
Train data
Test data
Data Preprocessing
Raw Data (csv file)

Software requirements:

- Anaconda navigator
- Jupyter Notebook
- Spyder

The following python libraries need to be installed:

- Numpy
- Pandas
- Scikit-learn
- Matplotlib and Seaborn
- Flask

4. Experimental Investigations

The purpose of this project is to predict the number of orders received by the food delivery service in a week.

1. Selecting the Dataset

   Machine Learning is dependent heavily on data which is needed for the “AI” to learn and thus it is the most crucial aspect that makes algorithm training possible. It was crucial to find a dataset that covers all the possible factors that affect this prediction so that it can be as accurate as possible.

   The dataset consists of three individual datasheets:

- train.csv (training data set):

  It contains the historical demand data for all centers,
  Contains information like id, week, center id, meal id, checkout price, base price, emailer for promotion, homepage featured, number of orders.
  This file is used for training.

The Train dataset consists of 9 variables and records of 423727 unique orders

- test.csv (testing data set)

  It contains information like id, week, center id, meal id, checkout price, base price, emailer for promotion, homepage featured. This file is used for testing. It contains the same features as the training dataset except for the target variable which is the number of orders. The Test dataset consists of 8 variables and records of 32573 unique orders.

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	The base price of the meal
emailer_for_promotion	Emailer sent for promotion of meal
homepage_featured	The meal featured at the homepage

num_orders	(Target) Orders Count

- Fulfilment_center_info.csv:

  contains the information of each fulfillment center. The dataset consists of 5 variables and records of 77 unique fulfillment centers.

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)

- meal_info.csv: contains the meal information.

Variable	Definition
meal_id	Unique ID for the meal

category	Type of meal 1. Beverages 2. Biryani 3. Desert 4. Extras 5. Fish 6. Other Snacks 7. Pasta 8. Pizza 9. Rice Bowl 10. Salad 11. Sandwich 12. Seafood 13. Soup 14. Starters
cuisine	Meal cuisine 1. Continental 2. Indian 3. Italian 4. Thai

2. Importing Python Libraries

1. Pandas: was imported for data manipulation.

2. NumPy: is used for numerical analysis.

3. Matplotlib and Seaborn: Both are data visualization libraries and are used for plotting graphs which will help us for understanding the data.

4. Pickle: was imported to serialize your machine learning algorithms and save the serialized format to a file.

3. Exploratory Data Analysis

It is an approach to analyzing data sets to summarize their main characteristics.

No null values were found.

All datasets were merged using the primary feature which is centre id

No null values were found in the merged dataset.

4. Label Encoding

Our dataset contained a lot of categorical variables which couldn't be processed by the AI. To solve this problem the Label Encoding technique was used which converted them into unique integer values which the machine was able to process in mathematical equations. To implement label encoding in Python I used the scikit-learn library.

5. Data Visualisation

This data needed to be visualized for which I used the Matplotlib and searborn library.

A line graph was used to display the relationship between the number of buyers and the number of orders as it is a good way to show a trend in the relationship between two values.

To show the correlation between different values a searborn heatmap was used.

6. Evaluation Metric

Various regression algorithm models were imported and then tested using the fit method to find the Root Mean Square Error. This method was used to minimize large errors.

- Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

  RMSLE: 129.75092195572807

- Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is

where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

- Elastic Net is an extension of linear regression that adds regularization penalties to the loss function during training.

  RMSLE: 133.28237995203088

- Decision Tree is a supervised machine learning model used to predict a target by learning decision rules from features.

  RMSLE: 62.86040686551593

- KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighborhood.

  RMSLE: 67.0011024616613

- Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.

  RMSLE: 67.0011024616613

- XGBoost is a powerful approach for building supervised regression models. Ensemble learning involves training and combining individual models (known as base learners) to get a single prediction, and XGBoost is one of the ensemble learning methods. RMSLE: 70.11319918873484

The ‘Decision Tree Model’ was used as it had the best (lowest) RMSE value of 62.86040686551593

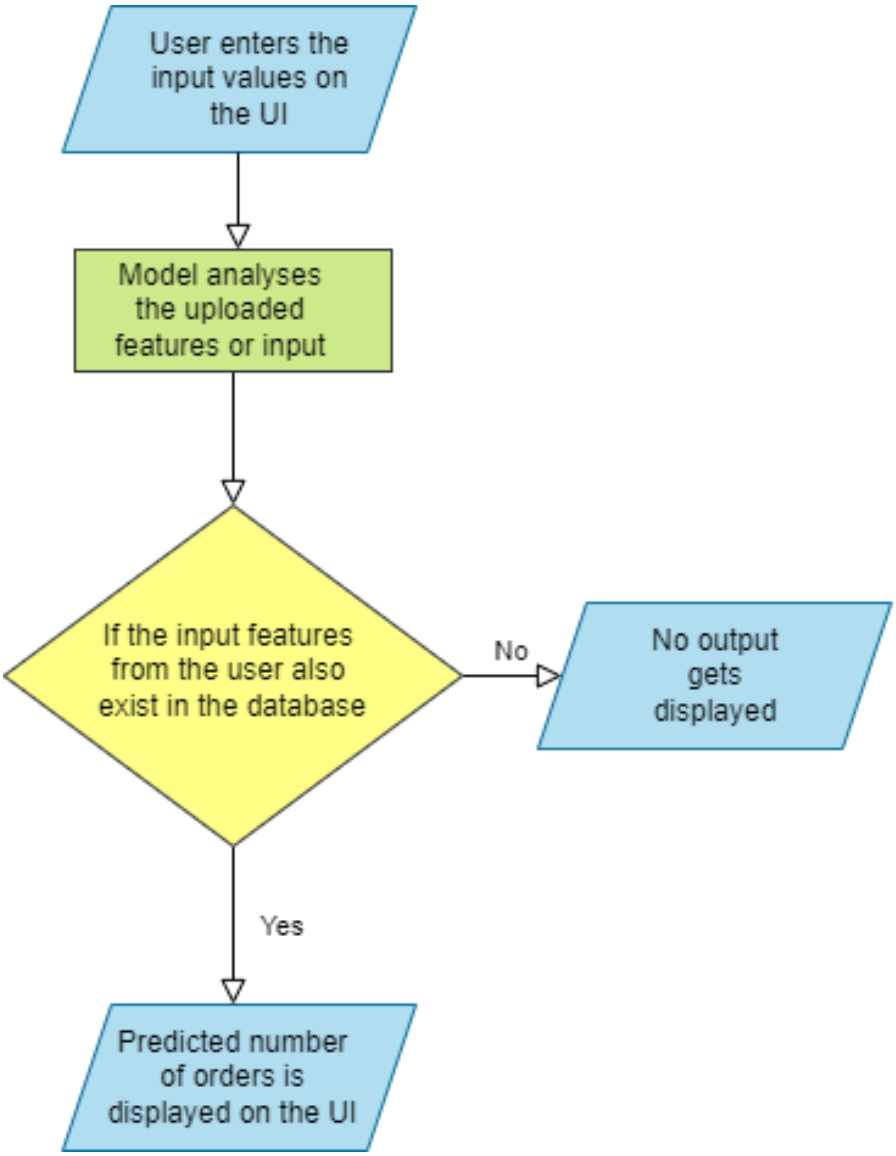
Testing data was used to test the model and the result was saved.

7. Splitting the Dataset

1. Independent variable (x): 'homepage_featured', 'emailer_for_promotion', 'op_area', 'cuisine', 'city_code', 'region_code', 'category' columns. These are the input variables.

2. The dependent variable (y): 'num_orders' column is considered as dependent variable.This is the output variable

5. Flowchart



The flow chart shows us the control flow of the solution. The input provided by the user is analyzed by our machine learning model which gives us the predicted number of orders as a result. The only condition being that the input values given by the users exist in our database.

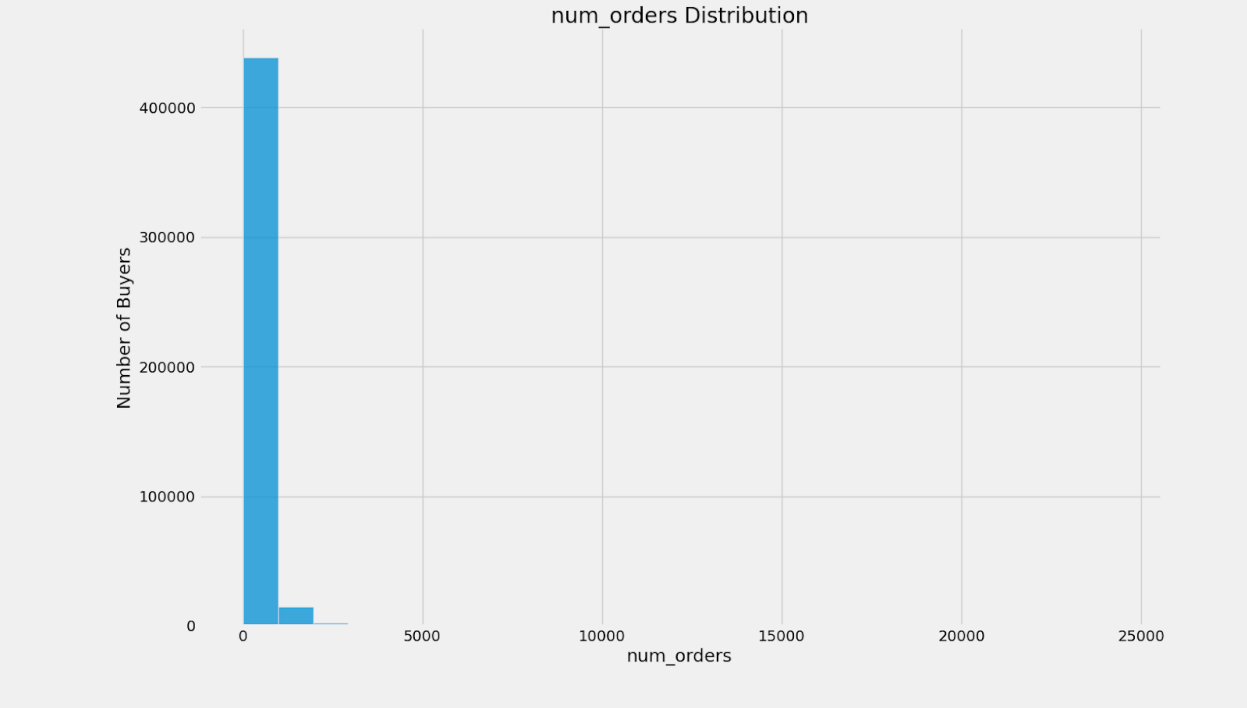
6. Result

The result was the number of orders in a given week when all the other factors are given as input by the user.

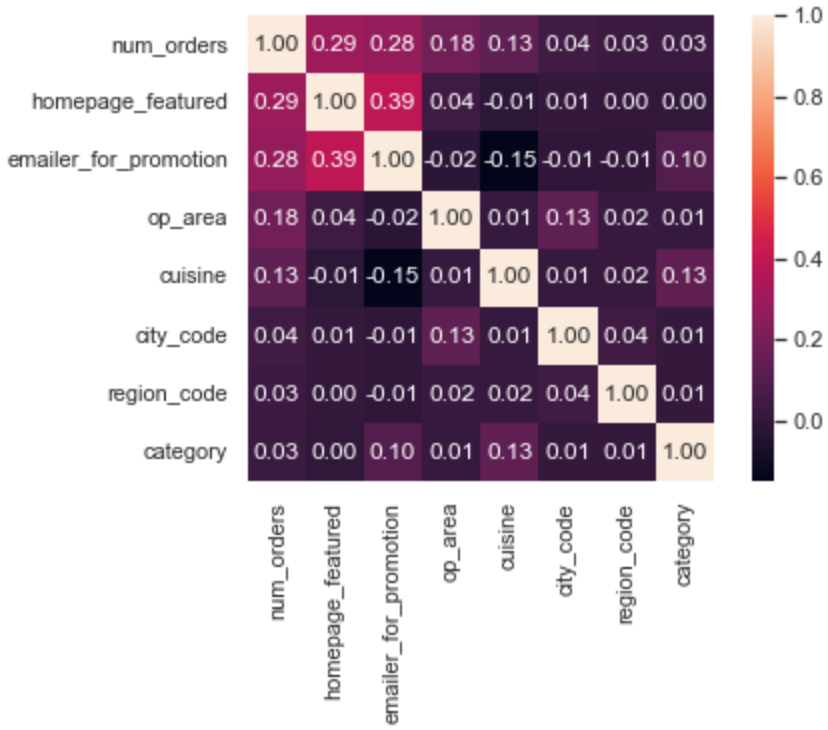
The output values obtained for the testing dataset are given in this link: [submission]

Data Visualisation:

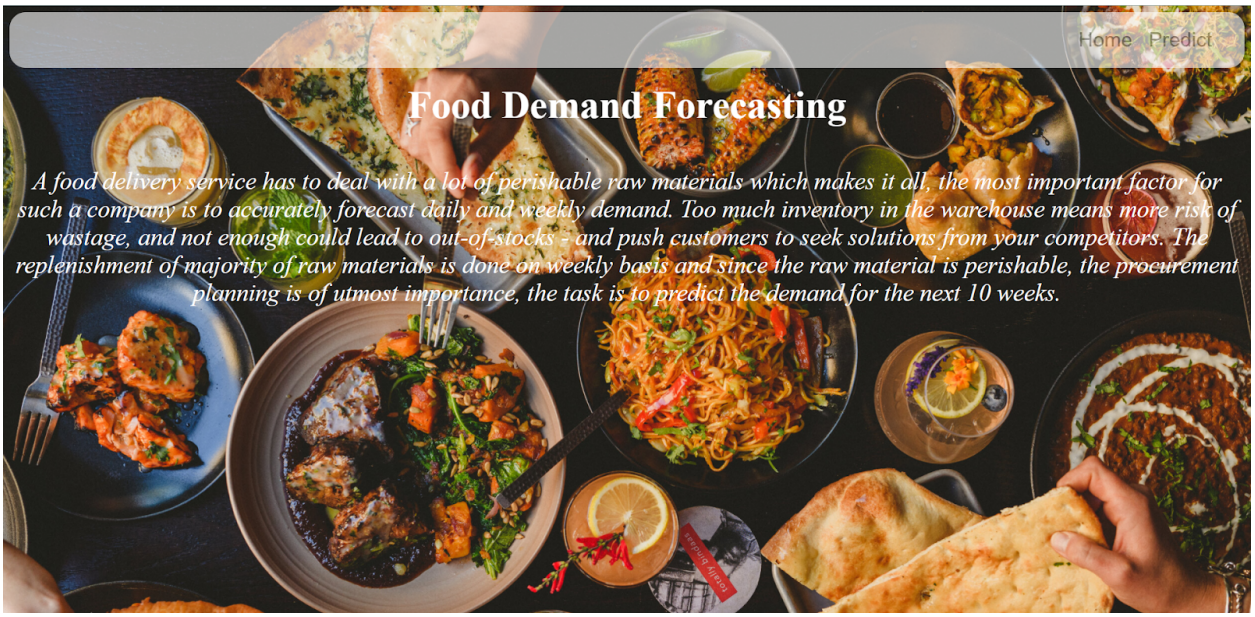
1. Bar Graph



2. Correlation Heatmap



3. Web Application



7. Advantages and Disadvantages

Advantages:

- Increased Supply Chain Efficiency

  By forecasting the sales and their time frame production, warehousing, shipping can be scheduled better. This helps you plan scheduled maintenance shutdowns away from busy sales periods and have adequate materials and labor on hand throughout the year. You can increase your supply accordingly when you are aware of the fluctuations in demand

- Improvement in Labor Management

  Having too few workers to handle a spike in sales orders can lead to slow order fulfillment which can lead to losing out on a loyal customer base. Having too many workers idle wastes money and resources.

- Ensure Adequate Cash Flow

  Knowing the peaks and valleys of demand helps you better manage your

cash flow, ensuring you have enough money on hand to pay vendors and suppliers on time. If they are not paid on time they may cut you off and this will hamper your production cycle. Demand forecasting lets you reserve cash or negotiate bridge loans or credit terms in advance.

- Create Accurate Budgeting

  You can allocate your resources wisely depending on the fluctuations in demand eg: more spending on advertising during a slow period.

Disadvantages of Demand Forecasting

- Scope for Uncertainty

  Not all the factors that affect the food delivery service can be predicted. For example, a severe weather event, consumers taste and preferences, market trends

  There is no guarantee that consumers will behave in the same consistent manner.

- Keep It Simple

  Many businesses, especially smaller ones, believe that demand planning is far too complex. They do not have the manpower nor the time to invest in creating a viable demand planning forecast.

8. Applications

1. Supply chain management: includes the movement and storage of raw materials, work-in-process inventory, and finished goods from point of origin to point of consumption.

2. Economic forecasting: is the process of making predictions about the economy

3. Earthquake Forecasting: tells us the time, location, and magnitude of future earthquakes so the damage can be minimized

4. Land Use Forecasting: undertakes to project the distribution and intensity of trip generating activities in the urban area

6. Player & Team Performance in Sports: PECOTA, is a sabermetric system for forecasting Major League Baseball player performance

7. Political Forecasting • aims at predicting the outcome of elections

8. Transportation Forecasting: the process of estimating the number of vehicles or people that will use a specific transportation facility in the future

9. Telecommunications Forecasting: Telecommunications service providers perform forecasting calculations to assist them in planning their networks

10. Product Forecasting: is the science of predicting the degree of success a new product will enjoy in the marketplace so investors know if it's worth the investment

11. Technology Forecasting: attempts to predict the future characteristics of useful technological machines, procedures, or techniques

13. Weather Forecasting: is the application of science and technology to predict the state of the atmosphere for a given location.

14. Flood Forecasting: the use of real-time precipitation and streamflow data in rainfall-runoff and streamflow routing models to forecast flow rates and water levels for periods ranging from a few hours to days ahead, depending on the size of the watershed or river basin.

9. Conclusion

By creating a system that predicts food orders with the help of a Machine Learning algorithm that provides more accurate data, a better understanding of food order patterns.
This is beneficial for making better business decisions, helping to solve the hunger crisis faced by the poor by reducing wastage and improving the ecological balance.
In this project, many steps have been taken to process the data and evaluate the accuracy of the algorithm so that overall the predictions made by the ML model are as accurate as possible.
The web application made using the flask framework gives the user a convenient and accessible interface to enter input values and get the prediction.
This project is a very good example of how we can use technology and artificial intelligence to make such a large positive impact.

10. Future Scope

I will use various advanced regression models like CatBoost and LightGBM Regressors to achieve a lower Root Mean Square Error Value. I will try different ensemble techniques.
Thus the predictions achieved will have higher accuracy.
I will use more data visualization techniques and thus make more aspects of the data understandable and presentable. I will use Business Intelligence tools like Power Bi and Google studio to further enhance the representation of the dataset.
Features and values of fulfillment centers, categories, regions, meal ids will be represented and analysed.

11. Bibliography

https://www.slideshare.net/bkdeepam/forecasting-introduction-its-applications

https://github.com/SaiPrasath-S/DemandPrediction

https://www.shipbob.com/blog/demand-forecasting/

https://medium.com/@roman.vinayagam24/food-delivery-demand-forecasting-2cabd23b9120

https://www.kaggle.com/kannanaikkal/food-demand-forecasting/code

Appendix

1. Source Code
   https://github.com/smartinternz02/SBSPS-Challenge-6507-Food-Demand-Forecasting-for-Food-Delivery-Company-using-IBM-Cloud