REPORT

GENPACT

(FOOD DEMAND FORECASTING)

NAME- Dev Chauhan

UNIVERSITY ROLL NO- 2013648

Problem Statement

Your client is a meal delivery company which operates in multiple cities. They have various fulfillment centers in these cities for dispatching meal orders to their customers. The client wants you to help these centers with demand forecasting for upcoming weeks so that these centers will plan the stock of raw materials accordingly.

Introduction

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. Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful. Given the following information, the task is to predict the demand for the next 10 weeks (Weeks: 146-155) for the center-meal combinations in the test set:

Historical data of demand for a product-center combination (Weeks: 1 to 145)

Product(Meal) features such as category, sub-category, current price and discount

Information for fulfillment center like center area, city information etc.

Motivation

1. The core motive for making this project was primarily an extreme interest towards an interesting area of research. Also food demand analysis is a hot topic for research.
2. Use of electronic media is increasing day by day.
3. Time is money or even more valuable than money therefore forcasting the demand of food is will save both time and money

TOOLS AND LIBRABIBS

1. **Pandas:** It is a library available for python programming generally used for data analysis or while dealing with data.

2. **Matplotlib:** It is a library available for python programming used to visualize the data such as to plot the graphs, bar charts, etc.

3. **NumPy:** It is a library available for python programming used for dealing with matrices and arrays and to apply mathematical operations and functions on our dataset in the present in the form of an array.

4. **Seaborn:** Seaborn is a library for making statistical graphics in Python. It builds on top of Matplotlib and integrates closely with pandas data structures.
5. **Math:** This module provides access to the mathematical functions defined by the C standard.

METHODOLOGY

First of all we have the data set containing following Information

Weekly Demand data (train.csv):
Contains the historical demand data for all centers, test.csv contains all the following features except the target variable

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

fulfilment_center_info.csv:
Contains information for each fulfilment center

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 information for each meal being served

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

Now I imported the libraries discussed above > Then I read all the files using pandas library and stored them into dataframe of the respective name>Then I merged meal_info and fulfilment_center_info with test and train file>Then I checked null values of my train file, there were no null values so I moved on>Then by using info() method I checked all the object values present in my train file>Then by using dummy matrices I converted all the object type values to int or float type and merged them into my train file and dropped the original columns >since id of each row is unique so it doesn't affects the prediction so I removed it>Then I saw the relation between weeks and num order by using pyplot which gave me the graph of no. of orders per week>Then to condense the data of weeks I converted into month, year and quarter>Then I dropped the week column from my train file>Then we tried to find the outliers in various column by using boxplot>Then by analyzing pearson correlation values I dropped number region code column>then i stored dependent variable in y and independent in x>then I split my train dataset into two parts one for training purpose and other for testing to model in the ratio of 8:2>then I used randomforest regressor to train my data model and used r2 score to calculate the approx. value of the data model>Then I applied this data model on test file and recorded all the values of num order into a csv file.

Leaning Outcome

Demand forecasting helps businesses make informed decisions that affect everything from inventory planning to supply chain optimization. With customer expectations changing faster than ever, businesses need a method to accurately forecast demand.

If you’re looking for an ecommerce fulfillment solution to help you improve demand forecasting, learn more about how ShipBob helps you replenish stock and deliver the experience that customers want. Request a pricing quote of our fulfillment services below.