Demand Forecasting for Food Analysis

Parth Sanghavi, Sushopti Gawade, Darshan Oswal, Ronish Rathod,
(Professor in Computer Technology Department ,Pillai College Of Engineering,New Panvel)

ABSTRACT :

Demand prediction is the process by which historical data is used to estimate the number of product customers to be purchased. An important aspect of a successful demand system is accurate prediction, due in part to the operational decisions made based on the results of predictability models. Using proposed algorithms such as Linear Regression, LASSO significantly improves predictive performance. The current task consists of using methods to predict the demand for food industry products, which direct its sales in the food service market, to support short- to medium-term production planning.

Keywords: Demand forecasting, food product, Lasso regression, Linear regression

I. INTRODUCTION

Demand forecasting is a key component to every growing online business. Without proper demand forecasting processes in place, it can be nearly impossible to have the right amount of stock on hand at any given time. A food delivery service has to deal with a lot of perishable raw materials which makes it all the more important for such a company 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. In this challenge, get a taste of demand forecasting challenge using a real dataset.

Customer Segmentation is the process of division of the customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. The technique of customer segmentation is dependent on several key differentiators that divide customers into groups to be targeted. Data related to demographics, geography, economic status as well as behavioral patterns play a crucial role in determining the company direction towards addressing the various segments

1)Weekly Demand Data

SR No	Name of the Parameter	Information about the parameter
1	Id	Unique Id
2	Week	Week Number
3	Center_id	Unique Id for fulfillment Center
4	Meal_id	Unique Id for Meal
5	Checkout_price	Final price including discount,taxes & delivery charges
6	Base_price	Base price of the meal
7	Emailer_for_Promotion	Mailer sent for promotion of Meal
8	Homepage_featured	Meal Featured at home page
9	num_orders	Orders Count(Target)

2)Fulfillment_center_info:

Name of the Parameter	Information about the Parameter
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

3)Meal_info:

Name of the Parameter	Information about the Parameter
Meal_id	Unique Id for Meal
Category	Types of Meals
Cuisine	Meal Cuisine

Performance Parameters:

SR No	Name of the Parameter	Information about the parameter
1	Week	Week Number
2	Category	Types of Meals
3	Cuisine	Meal Cuisine
4	Checkout_price	Final price including discount,taxes & delivery charges
6	Base_price	Base price of the meal
7	Emailer_for_Promotion	Emailer sent for promotion of Meal
8	Homepage_featured	Meal Featured at home page
9	num_orders	Orders Count(Target)
10	city_code	Unique code for City

II. Literature Survey

1.) In paper [1] the author is using the Autoregressive integrated moving average(ARIMA) model.ARIMA and Holt-Winters, as well as the performance metrics of accuracy of demand forecasting, which involve only the MAPE and U-Theil.

2.) In Paper [2] the author is using Exponential Smoothing Models the main exponential smoothing methods will be presented: The simple exponential smoothing method, the Holt's method and the Holt-Winters method are also used.

3.) Production planning and control for remanufacturing

In the paper the author uses, the number of clients is predicted using machine readings and a statistical analysis method with internal and external data in the universe.

4.) Food demand prediction using ML

In this study, the number of customers is predicted using machine learning method and statistical analysis method with internal and external data in the universe. The Bayesian Linear Regression, the Boosted Decision Tree Regression, and the Decision Forest Regression are used for machine learning, the Stepwise method is used for mathematical analysis methods..

Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

5.) Collaborative forecasting in the food supply chain: Conceptual framework

In the proposed framework, we focus on collaborative prediction of manufacturers and retailers. Through a systematic review of the literature, we have identified trends, gaps and areas for future research that include partnerships, information sharing and the forecasting process in the supply chain.

6.) A retail demand forecasting model based on data mining techniques

The proposed operating method combines the same storage areas according to their sales behavior using a bipartite graphical collection. Subsequently, a mixed prediction section combining a mid-range model and a Bayesian Network machine learning algorithm is applied.

7.) Seek to predict in restaurants using machine learning and statistical analysis

Bayesian Linear Regression, Advanced Decision Tree Depression, and Decrease Forest Decision is used in machine learning, Stepwise method is used for mathematical analysis method.

III. METHODOLOGY

The model focuses on customer segmentation based on several parameters like income, age, spending patterns, etc. The first step is to import the dataset. Then Analyzing and visualizing the data. With the identification of customers, companies can release products and services to that target customers.

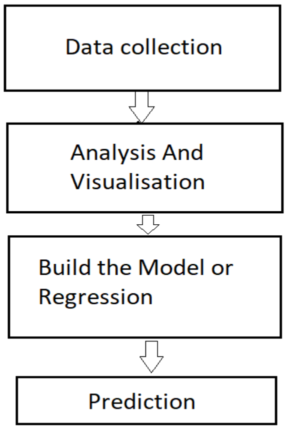

Fig.1 Proposed system Architecture

IV.Algorithm Details

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). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.
The acronym "LASSO" stands for Least Absolute Shrinkage and Selection Operator.

L1 Regularization

Lasso regression performs L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients. This type of regularization can result in sparse models with few coefficients; Some coefficients can become zero and eliminated from the model. Larger penalties result in coefficient values closer to zero, which is the ideal for producing simpler models. On the other hand, L2 regularization (e.g. Ridge regression) doesn't result in elimination of coefficients or sparse models. This makes the Lasso far easier to interpret than the Ridge.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

Performing the Regression

Lasso solutions are quadratic programming problems, which are best solved with software (like Matlab). The goal of the algorithm is to minimize:

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Which is the same as minimizing the sum of squares with constraint Σ |Bj≤ s (Σ = summation notation). Some of the βs are shrunk to exactly zero, resulting in a regression model that's easier to interpret.

A tuning parameter, λ controls the strength of the L1 penalty. λ is basically the amount of shrinkage:

- When λ = 0, no parameters are eliminated. The estimate is equal to the one found with linear regression.
- As λ increases, more and more coefficients are set to zero and eliminated (theoretically, when λ = ∞, all coefficients are eliminated).
- As λ increases, bias increases.
- As λ decreases, variance increases

Linear regression is one of the most commonly used predictive modelling techniques. The aim of linear regression is to find a mathematical equation for a continuous response variable Y as a function of one or more X variable(s). So that you can use this regression model to predict the Y when only the X is known.

This mathematical equation can be generalised as follows

$Y=\beta 1+\beta 2X+\epsilon$

Collectively, they are called regression coefficients and ? is the error term, the part of Y the regression model is unable to explain.

V. IMPLEMENTATION

```
library(readxl)
library(dplyr)
library(caTools)
library(mltools)
library(MASS)
library(leaps)
library(glmnet)
library(ISLR)
library(car)
train = read.csv("C:/Users/sonal/Desktop/food demand/train.csv")
test = read.csv('C:/Users/sonal/Desktop/food demand/test.csv')
meal = read.csv('C:/Users/sonal/Desktop/food demand/meal_info.csv')
fulfillment = read.csv('C:/Users/sonal/Desktop/food demand/fulfilment_center_info.csv')

--------------------------------Data Preprocessing--------------------------------

num_orders = 0
test = cbind(test, num_orders)
data = rbind(train, test)
data = merge(data, meal, by = 'meal_id')
data = merge(data, fulfillment, by='center_id')
data = data[-c(1, 2)]
data = data %>% arrange(week)
sum(is.na(data))
str(data)
summary(data$category)
summary(data$cuisine)
summary(data$center_type)
boxplot(log1p(data$num_orders))
data$num_orders = log1p(data$num_orders)
data$checkout_price = scale(data$checkout_price)
data$base_price = scale(data$base_price)
data$emailer_for_promotion = as.factor(data$emailer_for_promotion)
data$homepage_featured = as.factor(data$homepage_featured)
str(data)
train_ = data[1:nrow(train),]
test_ = data[(nrow(train) + 1):nrow(data),]

-------------------Separating train into train and validation set--------------------

train_set = train_[train$week <= 140,]
val_set = train_[train_$week > 140, ]
X = model.matrix(num_orders ~ .-week-id, data = train_set)[,-1]
X_val = model.matrix(num_orders ~ .-week-id, data = val_set)[,-1]
```

```
X_val = model.matrix(num_orders ~ .-week-id, data = val_set)[,-1]

-----------------------------Normal Linear Regression-------------------------------

lm.fit = lm(num_orders ~ .-week-id, data = train_set)
summary(lm.fit)
pred_lm = predict(lm.fit, newx = X_val)
sqrt(mean((val_set$num_orders - pred_lm)^2))

-------------------------------------------Subsetting-----------------------------------

nvmax_ = 10
subset.fit = regsubsets(num_orders ~ .-week-id, data = train_set, method='backward', nvmax = nvmax_)
k = summary(subset.fit)
k$adjr2
val.errors = rep(NA, nvmax_)
for(i in 1:nvmax_){
  coefi = coef(subset.fit, id=i)
  pred = cbind(1, X_val[,names(coefi[-c(1)])]) %*% coefi
  val.errors[i]= sqrt(mean((val_set$num_orders - pred) ^2))
}
val.errors
val.errors[which.min(val.errors)]

------------------------------------Lasso Regression--------------------------------------

ld = 10^seq(10, -2, length=150)
y = train_set$num_orders
rig.fit = glmnet(X, y, alpha=1, lambda=ld)
plot(rig.fit)
cvglm = cv.glmnet(X, y, alpha=1, lambda=ld, nfolds=10)
plot(cvglm)
best = cvglm$lambda.min
best
predict(rig.fit, s=best, type='coefficients')
pred_rl = predict(rig.fit, s=best, newx = X_val)
sqrt(mean((val_set$num_orders - pred_rl)^2))
test_set = model.matrix(num_orders ~ .-week-id, data = test_)[,-1]
pred_test = predict(rig.fit, s=best, newx = test_set)
submission = data.frame('id'=test$id,
                        'num_orders'=exp(pred_test))
names(submission)[2] = 'num_orders'
write.csv(submission, 'RL.csv', row.names = F)
-----------------------------------END----------------------------------------
```

VI. Result And Discussion

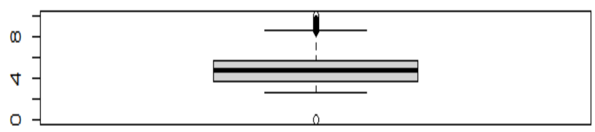

Fig-1

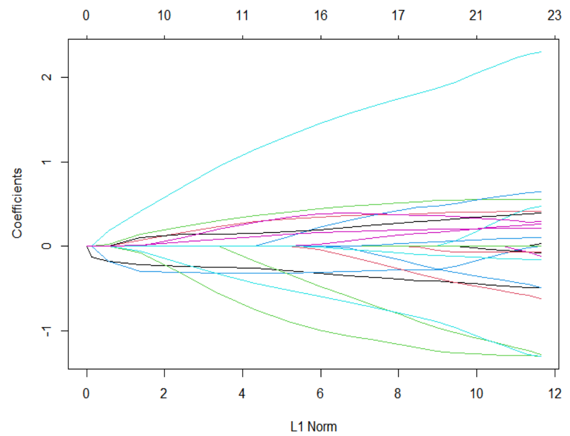

Fig-2

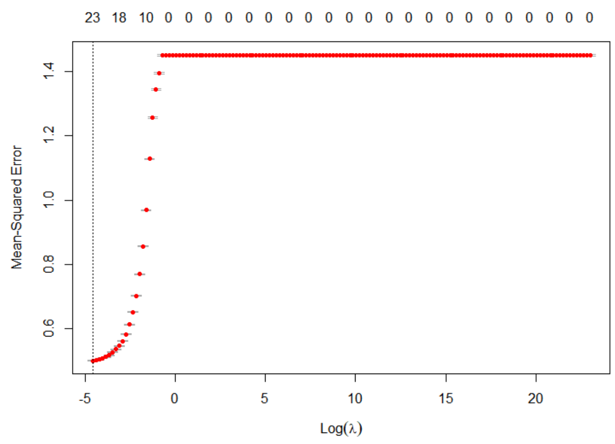

Fig -3

VII. Conclusion

Thus we have successfully Implemented Demand Forecasting for Food analysis which will help the businessmen to have an analysis of what type of a food has more demand in a specific area.This will also help in reducing wastage of food and serving fresh Food.

VIII. References

References : We have referred below links for the project :

1. https://github.com/RutvijBhutaiya/Food-Demand-Forecasting-Challenge
2. https://github.com/yashkim77/AV_Food_Demand_Challen
3. 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.
4. K.Aishwarya1, Aishwarya.N.Rao2, Nikita Kumari3, Akshit Mishra4, Mrs.Rashmi M R5June 2020 'FOOD DEMAND PREDICTION USING MACHINE LEARNING',Volume: 07 Issue: 06
5. Eksoz C., Mansouri S. A. and Bourlakis M. 2014 Collaborative forecasting in the food supply chain: Conceptual framework', International journal of production economics, Vol. 158, pp. 120-35.
6. İrem İşlek and Şule Gündüz Öğüdücü," A Retail Demand Forecasting Model Based on Data Mining Techniques".

ISSN: 2582-3930

7. https://www.researchgate.net/publication/329962590_Demand_forecasting_in_restaurants_using_machine_learning_and_statistical_analysis
8. https://statisticshowto.com/lasso/regression
9. 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.
10. https://towardsdatascience.com/supportvectormachines-svm-c9ef22815589