MINOR PROJECT

Final Evaluation Report

Automated Grocery List

Name of Students: Kapil Israni , Ayush Nagar , Akshara Nigam

Enrollment Nos: 17104011 , 17104012 , 17104018

Name of Supervisor: Mrs. Sharddha Porwal

Submitted in partial fulfilment of the degree of Bachelor of Technology in Information Technology.

Department of Computer Science Engineering & Information Technology.

Jaypee Institute Of Information Technology, Noida

Table of Contents

Page No.

Declaration

We hereby declare that this submission is our own work and that to the best of my knowledge and belief it contains more material previously published or written by another person no material which has been accepted for the reward of any degree or diploma of the university or any other Institute of higher learning, except where due acknowledgement has been made in the text.

Date:- 25/11/2019

Kapil Israni (17104011)

Ayush Nagar (17104012)

Akshara Nigam (170104018)

Certificate

This is to certify that the work entitled "**Automated Grocery List"** submitted by **Kapil Israni, Ayush Nagar** and **Akshara Nigam** of B.Tech (I.T) of Jaypee Institute of Information Technology Noida has been carried out under my supervision this work has not been submitted partially or wholly to any other university or institute for the award of any other degree or diploma.

Signature of the Supervisor

Name of the Supervisor : Mrs. Shraddha Porwal

Date : 22/11/2019

Acknowledgement

First and foremost we would like to thank our guide Mrs. Shraddha Porwal of Jaypee Institute of Information Technology,Noida for guiding us thoughtfully and efficiently throughout this project, giving us an opportunity to work at our own pace along our own lines, while providing us with very useful directions whenever necessary.

We would also like to thank our friends and classmates for being great sources of motivation and for providing us encouragement throughout the length of this project. We offer our sincere thanks to other persons who knowingly or unknowingly helped us in this project.

Signature of Students

Kapil Israni (17104011)

Ayush Nagar (17104012)

Akshara Nigam (17104018)

Summary

Our project is a tool for shoppers, as it learns from their previous purchases and predicts the next item that should be purchased along with their quantity. It also recommends the user, items he/she could buy by referencing other similar users. It extracts the best features and removes the outliers to give the user the desired result.

List of Figures

List of Tables

1. Introduction

1.1 General Introduction

We often find ourselves in a situation where we tend to forget the items we have to buy or the quantity of the item needed to be purchased and the related stuff. Its then when we wish for an application that would predict for them what all items that they should buy, by seeing the user's past purchasing trends. Also people wish for the recommendation. For example, if one buys bread and milk, they would want an application to recommend them butter and jam too.

Here is when our project **"Automated Smart List"** plays its part.

1.2 Problem Statement

Through our project, we wish to solve the problem of customers by simply automating their buying cycle and analyzing it. The customer only needs to open the application, log in to it, and click on **automate** or **recommend** feature buttons. Automation would predict the grocery items and their names along with the predicted quantity that a user should buy, while recommendation would recommend the products the consumer could buy according to his past purchases.

1.3 Significance/Novelty of the Problem

Most of the applications are based on general shopping list checklists, while this application is completely automated and the user can also add his items to the list, change them, delete them according to their needs. Also, we take care of the outliers so that they do not hinder in the prediction process. The recommendation is done on the basis of a global dataset.

1.4 Brief Description of the Solution Approach

The main objective of the project is to create a progressive web application as a tool to analyse the consumer product buying interval and predict the items and their respective quantity from the dataset of their transactions and also to recommend items to buy.

The database consists of collections namely :

1. Transaction table: Stores the entire buying history of the user along with date and quantity of purchase
2. Customers Table: Stores the user id, username and password for authentication.
3. Category Table: Stores the product category name along with a category item.
4. Item List Table: Stores the product name along with unique product id.
5. RTA Table: Stores the average buying rate of an item of each customer for each and every item that he/she has purchased
6. Recent_Purchases Table: Stores the recent purchases of each user.

Algorithms used :

It uses the best of two algorithms **Random Forest Regression** and **Supports Vector Regression** to train the model on the basis of purchase frequency gap and quantity.

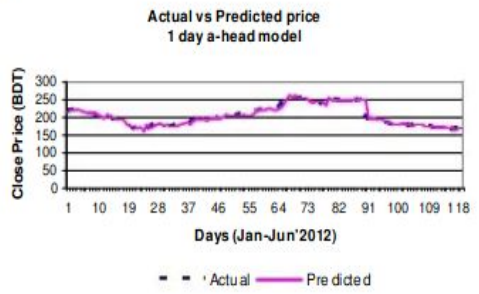
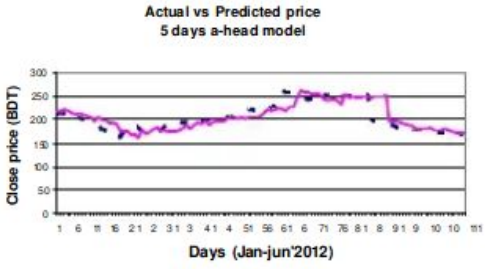
While Recommendation is done with the help of **Collaborative Filtering** using cosine similarity**.**

All the routing for the react framework is carried out from the front-end and communication between the two API's is done through JSON. The flask API that contains the python code is hosted on one platform while the react code is hosted on the other. As soon as the API is hit, the solution is generated and shown on the front-end.

2. Literature Survey

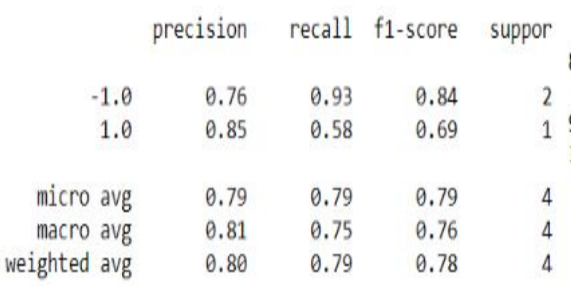
2.1 Paper 1

Title of the paper	Predicting Stock Market Price Using Support Vector Regression
Authors	Phayung Meesad and Risul Islam Rasel
Year of publication	2013
Publishing Details	International Conference on Informatics, Electronics and Vision (ICIEV)

Objective	Results
In this study, support vector regression (SVR) analysis is used as a machine learning technique in order to predict the stock market price as well as to predict stock market trend. Moreover, different types of windowing operators are used as data preprocess or input selection technique for SVR models. This is a new approach which uses different types of windowing functions as data preprocess for predicting time series data. Support vector regression is a useful and powerful machine learning technique to recognize pattern of time series dataset. It can produce good prediction result if the value of important parameters can be determined properly.	 

2.2 Paper 2

Title of the paper	Stock Market Prediction Using Machine Learning Algorithms
Authors	K. Hiba Sadia,Aditya Sharma, Adarrsh Paul, SarmisthaPadhi,Saurav Sanyal
Year of publication	4, April 2019
Publishing Details	International Journal of Engineering and Advanced Technology (IJEAT)

Objective	Results
The main objective of this paper is to find the best model to predict the value of the stock market. During the process Of considering various techniques and variables that must be taken into account, we found out that techniques like random forests, support vector machine were not exploited fully. In this paper we are going to present and review a more feasible method to predict the stock movement with higher accuracy	

	precision	recall	f1-score	suppor
-1.0	0.76	0.93	0.84	2
1.0	0.85	0.58	0.69	1
micro avg	0.79	0.79	0.79	4
macro avg	0.81	0.75	0.76	4
weighted avg	0.80	0.79	0.78	4

3. Requirement Analysis and Solution Approach

3.1 Functional and Non-Functional Requirements

1) **Support Vector Regression**: It uses the main idea: To minimize error, individualizing the hyperplane that maximizes the margin, keeping in mind that part of the error is tolerated.

2) **Random Forest Regression**: It is an ensemble for Decision Regression Trees and helps in classifying. It computes proximities between pairs of cases that can be used in clustering, locating outliers, or (by scaling) give interesting views of the data.

3) **Collaborative Filtering**: The motivation for collaborative filtering comes from the idea that people often get the best recommendations from someone with tastes similar to themselves. Collaborative filtering encompasses techniques for matching people with similar interests and making recommendations on this Collaborative filtering algorithms often require (1) users' active participation, (2) an easy way to represent users' interests, and (3) algorithms that are able to match people with similar interests.

4) **MongoDB**: It is a cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with the schema.

5) **ReactJS**: It is a Javascript frontend framework that creates virtual dom which reduces the loading time of a page. The division which has changed only gets reloaded and rest of the divisions remain unchanged.

3.2 Architecture

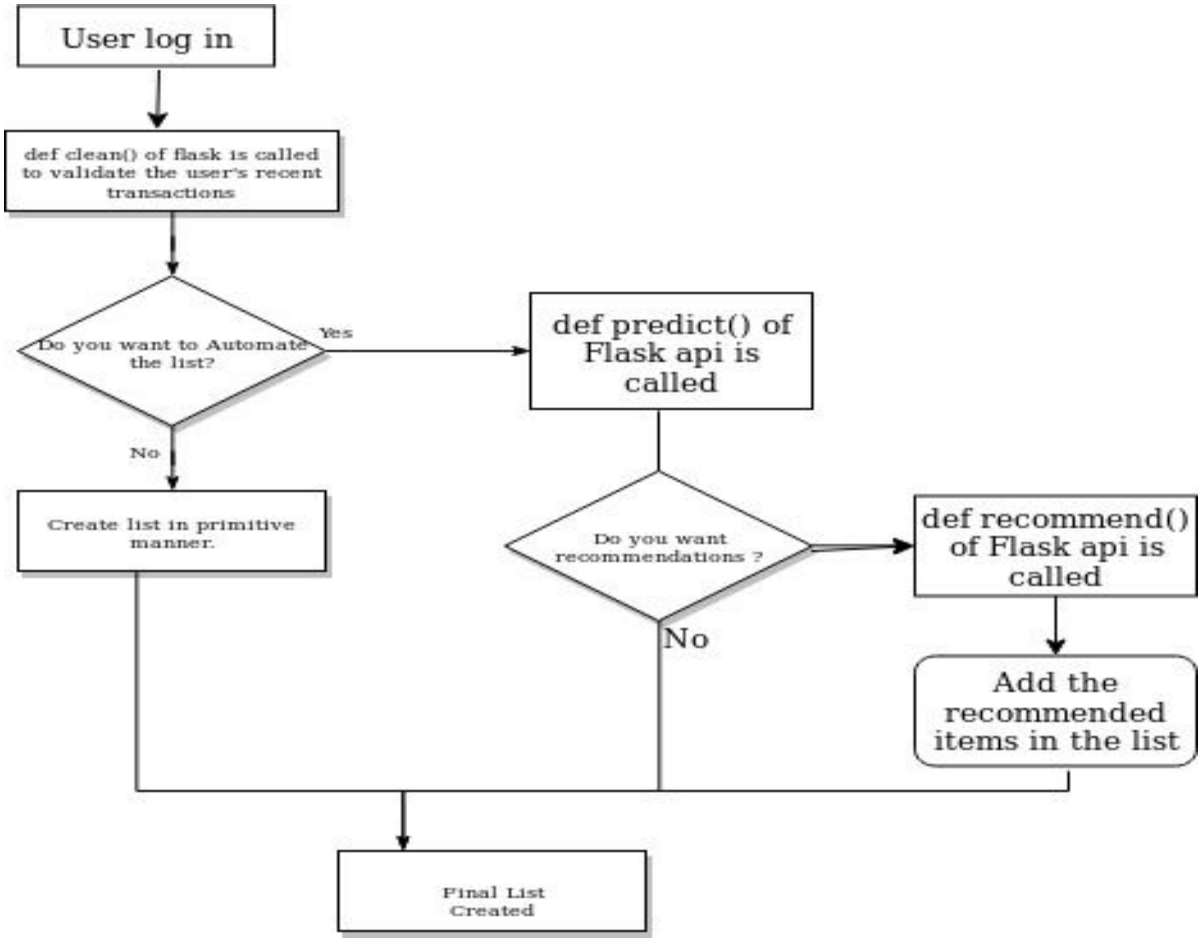


Fig 1: App working Diagram

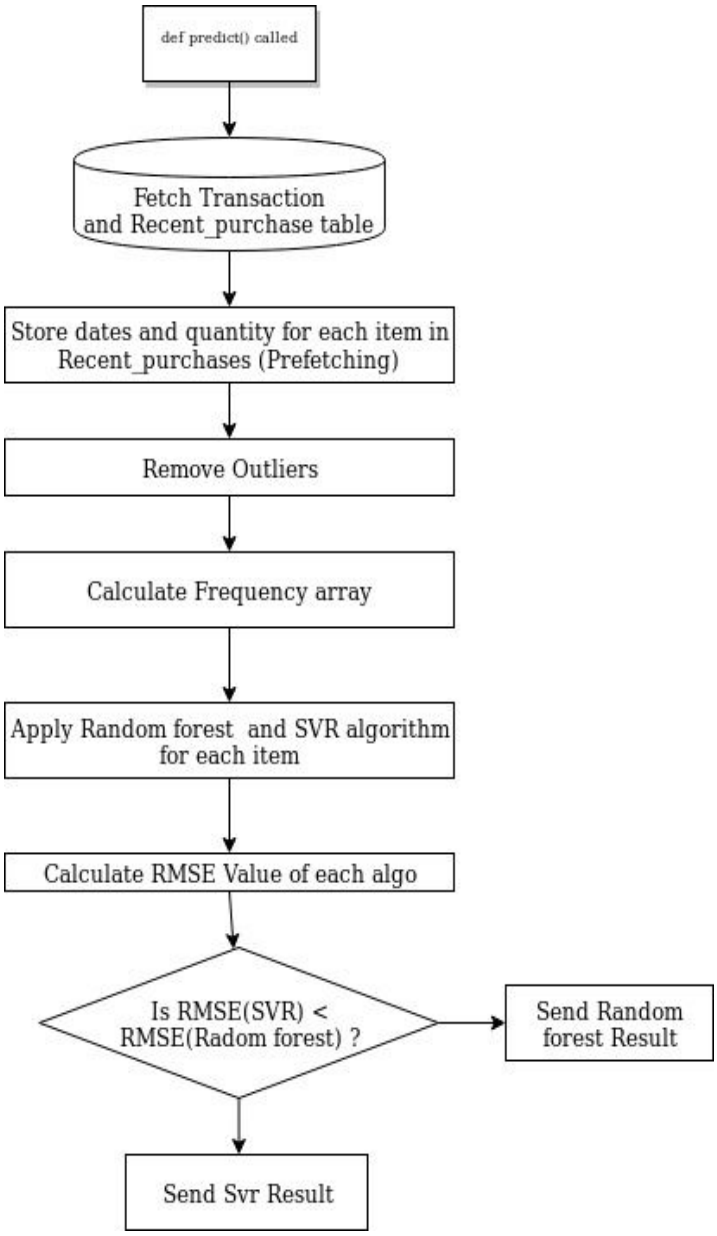


Fig 2 : Working of prediction Algorithm

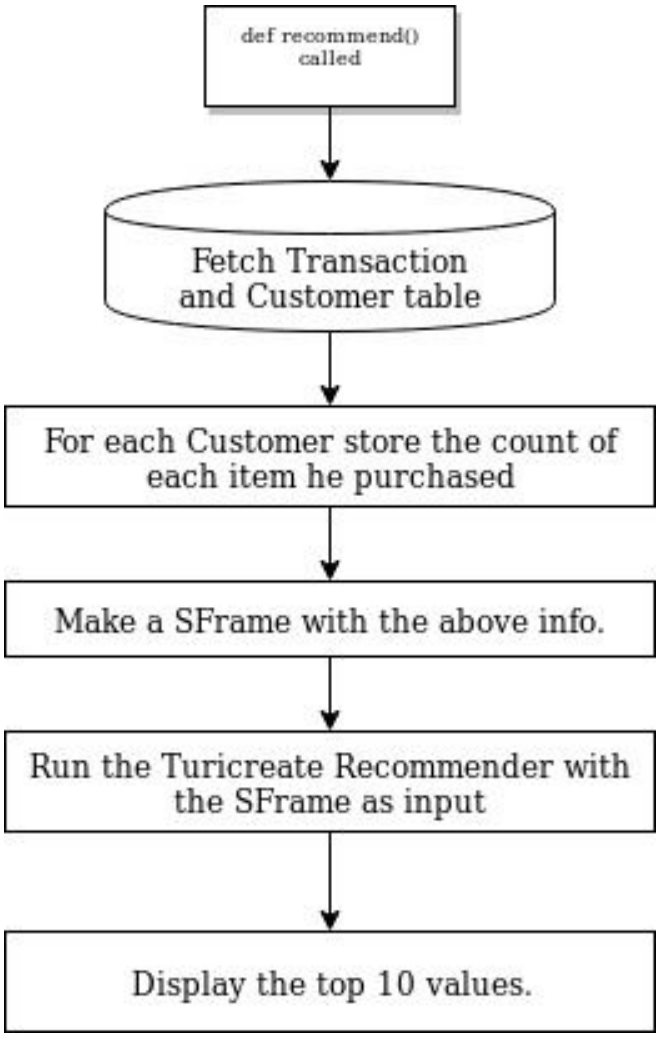


Fig 3: Working of Recommendation System

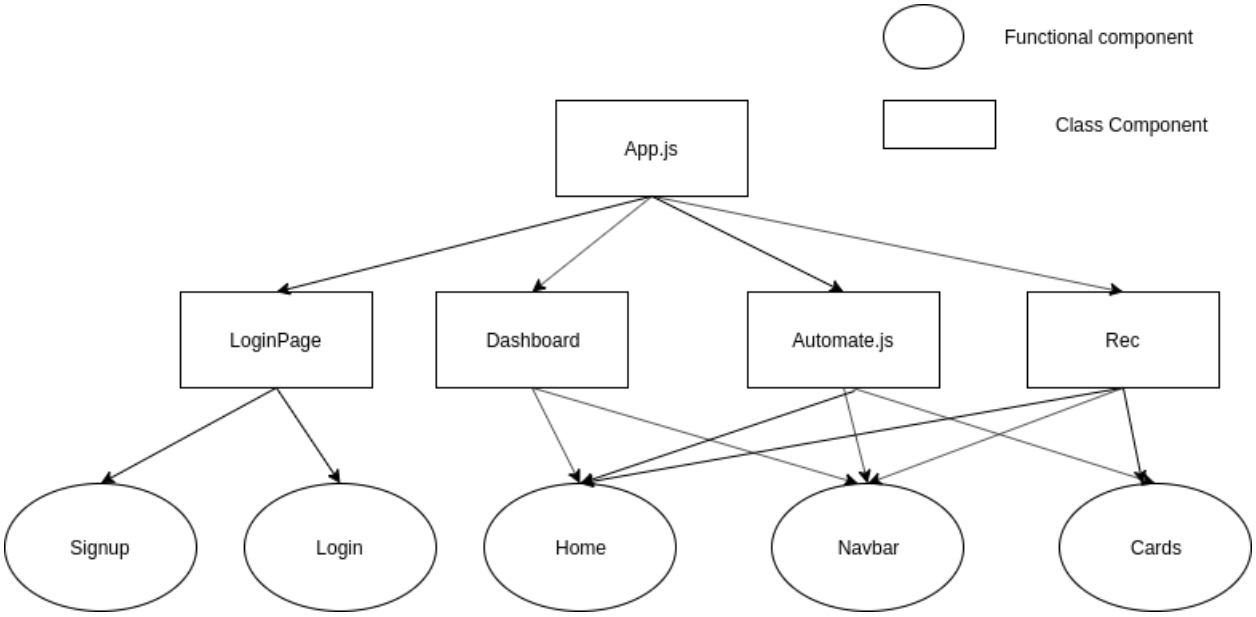


Fig 3: Component Diagram of React

3.3 Solution Approach

The working of the tool starts with :

Feature Extraction: Out of many features, the only Date of purchase, Quantity of purchase for a particular Customer Id and Item Id were selected. These dates were helpful to know the time gaps between the last purchases and hence train the model based on the time gaps.

Removing Outliers: When we work with a real dataset, we know that no dataset is perfect and always do contain the outliers. So we tend to remove the outliers using the Z score values with default threshold =3.

Testing and Training: Once the features are selected and outliers are removed, the model is fitted for random forest regression and support vector regression. Testing and prediction are done only one of the best algorithms based on their errors.

Front-end: Reactjs is used as a front-end framework. It is made up of multiple components. Components make the code modular and faster. As one login, the predict and recommend routes of flask APIs are called to hide the computation of both the algorithms.

Back-end:

A. Node.js is used for backend work. This helps the react and flask and React API to be separately used even in case of any api breakdown.
B. Flask API : The sole of this project is the Flask Api that handles requests and performs the following tasks
    a. Prediction : Returns the value of the items to be added in the list with their predicted outputs.
    b. Recommendation: Returns the recommended the items that a user should buy .
    c. Credential Verifications: Verifies the login requests of the user
    d. Clean up : Cleans up the Recent Purchases of the user that are no longer recent items anymore

The reason for a central API was to allow any device to access the information and add up as a member of this shopping list app.

3.4 Algorithm Details:

1) Support Vector Regression :(SVR) is used as a machine learning technique for pattern recognition. The training is basically done with the input as the Frequency of purchase (For a particular item) and train for the output as quantity.

   Kernel Information : The function used to map a lower dimensional data into a higher dimensional data.The kernel we are using here is the Radial Basis Function (RBF) Kernel.

   Model Details :

Model Name	Constant	Gamma
Radial Basis Function Kernel	1000	0.3

C = cost of misclassification
Gamma = Gaussian Kernel parameter

2) Random Forest Regression : It is an ensemble for multiple decision regression trees. In the CART algorithm, a decision tree is built by determining the splits of nodes that, lead to the greatest reduction in Gini Impurity, which means that the decision tree tries to form nodes containing samples of randomly chosen data from and finding values in the features that cleanly divide the data into classes.

   The Gini Impurity of a node is the probability that a randomly chosen sample in a node would be incorrectly labeled if it was labeled by the distribution of samples in the node.

   Random Forest uses two key concepts :

   a) Random sampling of training data points when building trees (Bagging)

   b) Random subsets of features considered when splitting nodes

The random forest combines hundreds of decision trees, trains each one on a slightly different set of the observations, splitting nodes based on the standard deviation and variance with respect to the output. The final predictions of the random forest is made by averaging the predictions of each tree.

3) Collaborative Filtering: This is a Item-Based Collaborative Filtering (IB-CF) i.e a Recommendation based on calculating similarities of two items based on Customers purchasing count of two items.

We have used a Cosine Similarity is a metric used to measure how similar the two items or documents are irrespective of their size. It measures the cosine of an angle between two vectors projected in multi-dimensional space.

On the basis of cosine score is calculated and hence the first 10 items with highest sore are recommended.

$$\text{Similarity}(p,q) = \cos\theta = \frac{p \cdot q}{\|p\|\|q\|} = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^{\,2}}\sqrt{\sum_{i=1}^{n} q_i^{\,2}}}$$

Formula of cosine similarity

4. Implementation

Implementation Details

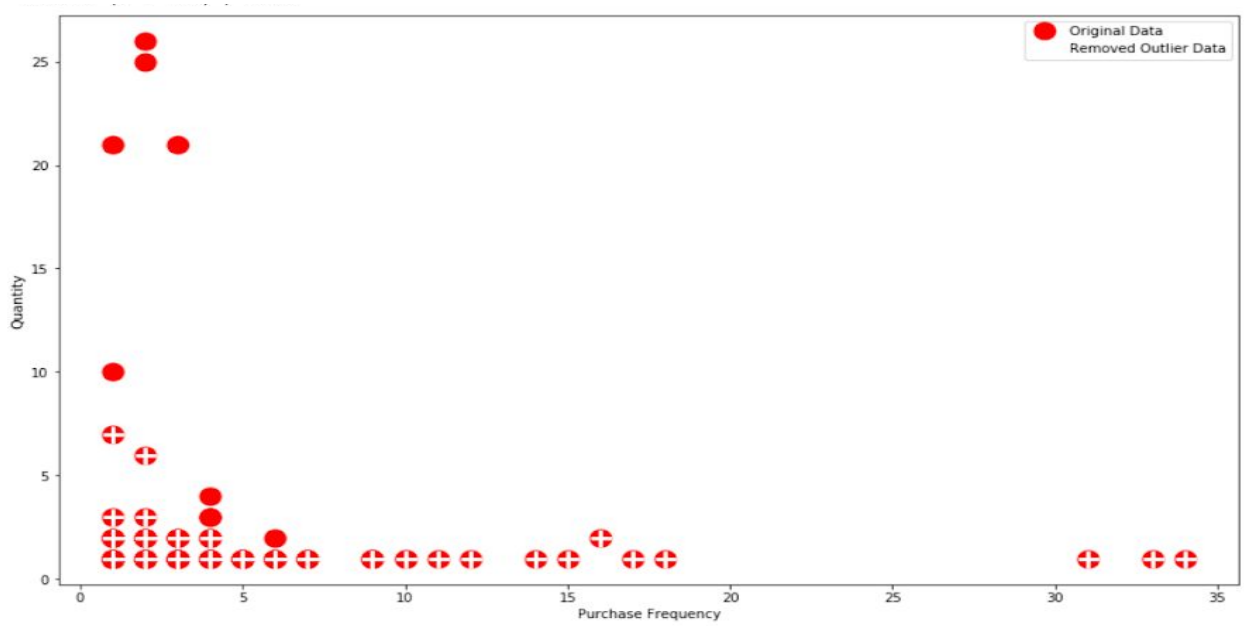


Fig 4: After removing outliers for cust_id 25 and item_id 2808

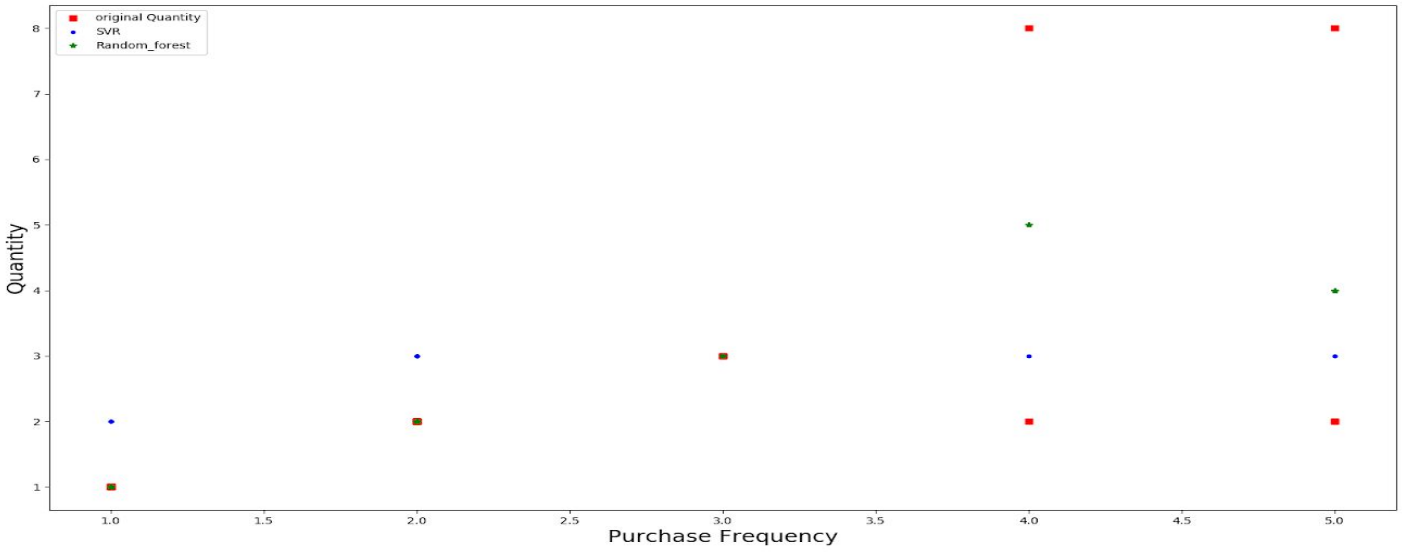


Fig 5 : Predicted values for the quantity from SVR and Random Forest Regression

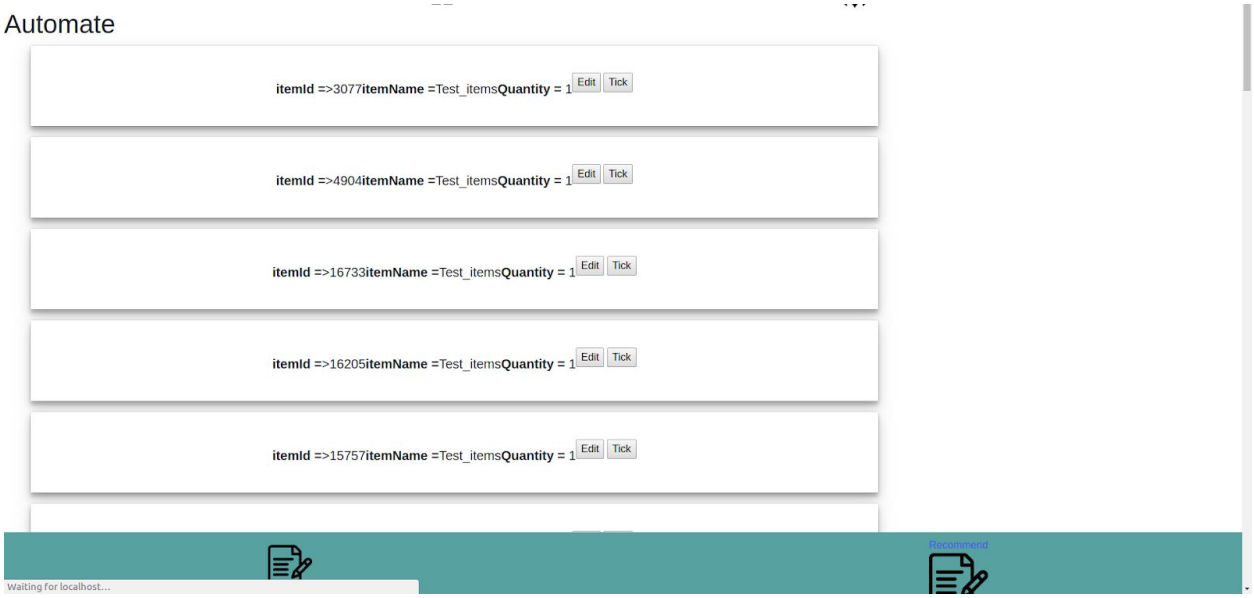


Fig 6: Predicted items on the frontend with rea

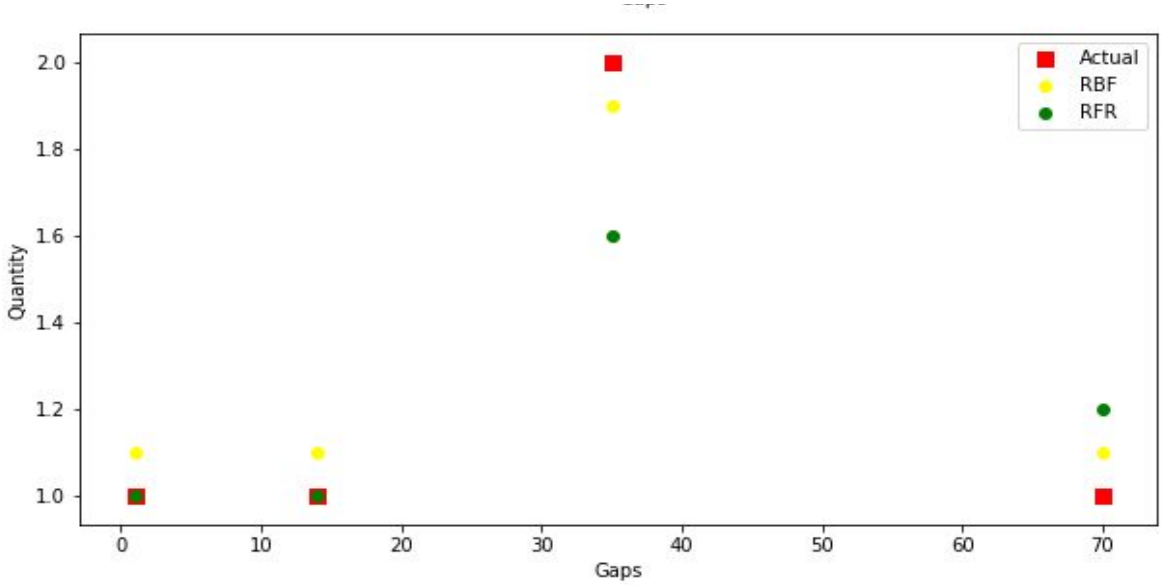


Fig 7 : Predictions Results

5. Testing

5.1 Unit Testing

The dataset was made fit to be trained by the model by removing outliers by using z-value, dropping the unnecessary columns, creating a new table rta, by finding the average of the purchases and optimizing the calculations.

5.2 Testing and Results

The project stood upon our expectations while successfully completing the following tasks :-

a) Hosting the application on Azure, thereafter using cloud computation.

b) Ignoring the outliers to avoid error in calculation.

c) Understanding the actual problem statement and using the dataset properly and creating a well-defined database.

Fig 8: Shows RMSE error for the customer ID 15

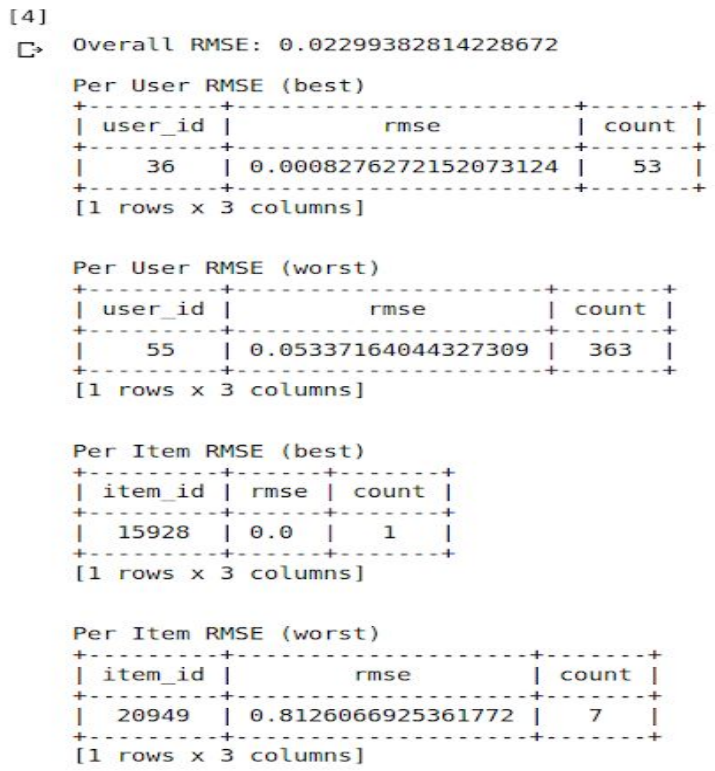


Fig 9: Shows the RMSE of the recommendation system both user and item wise.

6. Findings and Conclusion

6.1 Findings

The dataset used for the project was taken from kaggle. It comprises category.csv, itemlist.csv, sales_prediction.csv and customers.csv which was fed in our database. Also for our use, we created RTA and recent tables to avoid checking a huge dataset multiple times which would increase time consumption and make the PWA slow.

After passing the data to our model, we realized that the following features were important:-

a) date
b) quantity
c) item_id
d) category_id

6.2 Conclusion

The project stood upon our expectations while successfully completing the following tasks :-

d) Hosting the application on Azure, thereafter using cloud computation.

a) Ignoring the outliers to avoid error in calculation.

b) Understanding the actual problem statement and using the dataset properly and creating a well-defined database.

6.3 Future Scope

Though the project has stood up to our expectations yet there is always room for improvement that can be added to it.

a) Bill / QR scanning so that the items automatically get added to the list.

b) Giving the customer suggestions from where to buy looking at the previous shops he bought from.

References

[1] K. Hiba Sadia,Aditya Sharma, Adarsh Paul, SarmisthaPadhi, Saurav Sanyal, “Stock Market Prediction Using Machine Learning Algorithms”, International Journal of Engineering and Advanced Technology (IJEAT) - 4 April 2019

[2] Phayung Meesad and Risul Islam Rasel, “Predicting Stock Market Price Using Support Vector Regression”, International Conference on Informatics, Electronics and Vision (ICIEV) - 2013

[3] Mithileysh Sathiyanarayanan, “Data Analysis and Forecasting of Grocery Sales in Ecuador”- May 2011

[4] Bawa, Kapil, Ghosh, Avijit, “A Model of Household Grocery Shopping Behavior”, Kluwer Journal, May,1999

[5] Eric T.Bradlow, ManishGangwar, PraveenKopalle, SudhirVoleti, “The Role of Big Data and Predictive Analytics in Retailing”, Journal of Retailing - March 2017