**CLOUD TECHNOLOGY ASSIGNMENT 2**

****

**Final Report  
Intelligent Product Recommendation in Grocery Retail**

|  |  |
| --- | --- |
| **Project Title:** | Intelligent Product Recommendation in Grocery Retail |
| **Group Members:** | Aritra Dutta, Garima Singh, Kiran Negi, Parnab Das, Saumitra Das, Souradip Goswami, Yangchen Dolkar Sherpa |
| **Group coordinator:** | Aritra Dutta |
| **Student IDs:** | 19211293, 19211010,19210510, 19211289,19211286,19210273,19210377 |
| **Submission Date:** | December 15th, 2019 |

**1. INTRODUCTION:**

All ecommerce websites today focus widely on customer satisfaction and shopping experience; product recommendation plays an important role in generating revenue for ecommerce platforms. Product recommendations is a personalization strategy where products are populated dynamically to a user based on various user attributes which are likely to be purchased by a user. Our project aims at building such an application where we try to predict the likelihood of products to be purchased in a group. For this we have identified data from R datasets package named groceries and have processed the data as per our requirements using PIG. We have done our analysis on the processed data using Apriori algorithm and Association rule learning in R and then we have built an interactive shiny application where the data needs to be entered by the user and the application carries out the data analysis on the given data. Once the application is built, we put all the packages used by our application in a container and published the same in shinyapps.io platform to be used by everyone who intend to use the application.

**2. DATA COLLECTION, CLEANING AND PREPARATION:**

The data collection, cleaning and processing of the data is an integral part for developing this application, we have built an interactive application where the user can insert personal data and run the application to understand product recommendation summary based on the data they fed to the application.

**2.1 DATA COLLECTION**

To build our application, we have used the “groceries” dataset available in the “datasets” package in R. The data in this package is in the form of transactions, where each row is considered as an individual transaction.

**2.2 DATA CLEANING AND PREPARATION**

The “groceries” dataset from R includes each transaction as rows and it needs to be processed into two attributes which gives the transaction number and the item for that transaction. For example, if there are two items in a transaction as T1: {milk, coffee}, we needed to break it down as {T1, milk} and {T1, coffee}. We achieved this by using PIG Latin to clean the groceries dataset and process it as per our needs.

**3. DATA PROCESSING- APRIORI ALGORITHM**

Apriori algorithm is a machine learning technique which generates association rules that define the dependencies of two or more objects with a certain probability. Let us consider an example where a transaction has items {orange, mango, banana}, here an association rule is defined as an implication of the form where, {orange, mango} => {banana}, we use three prime metrics Support, Confidence and Lift to measure the precision of an association rule.

**3.1 SUPPORT**

The support metrics define the frequency of an item in the dataset and is denoted as below:  
 

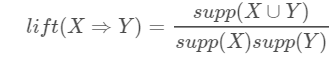
**3.2 CONFIDENCE**

A rule’s confidence is a measurement of its predictive power or accuracy, so for a given rule X => Y, the confidence can be denoted as below:



**3.3 LIFT**

The ratio of the observed support to the expected support of each items independently denotes the lift:



**4. RELATED WORK**

The three major components of our application are data mining, market basket analysis and Apriori algorithm. The paper on “Market basket analysis using Apriori algorithm in data mining” by Swati Mahesh Joshi laid the foundation for understanding the concepts required in designing the application. [1] We got the understanding of association rule mining in market basket analysis from the paper “A Survey on Association Rule Mining in Market Basket Analysis” by Savi Gupta and Roopa Mamtora. [2] Along with this the market basket analysis module on Datacamp and Udemy were also very helpful in understanding the low-level dependencies. “A Dynamic Application of Market Basket Analysis with R and Shiny in The Electric Materials Sector” by KARAHAN ADALI and M. Erdal BALABAN laid the foundation of developing a shiny interface for our application.[3]

**5. CHALLENGES AND LESSONS LEARNED:**

The major challenges in developing this application include gathering of proper data, processing the data for use by the application, building the application using association rules and deploying the application to a cloud platform. Interestingly all these challenges helped us learn various aspects of cloud technologies. Learning about data mining concepts using Apriori algorithm and deploying a live application to a cloud platform were the major highlights.

**6. RESPONSIBILITY STATEMENT:**

Each group member contributed equally to the development of the application. From the very beginning we understood that the majority work of this project is to self-learn and every team member acted responsibly.

|  |  |  |
| --- | --- | --- |
| **Group Member** | **Tasks** | **Member Contribution** |
| Souradip Goswami | Data collection, cleaning, processing | Satisfactory |
| Parnab Das | Data collection, cleaning, processing | Satisfactory |
| Saumitra Das | Apriori algorithm and shiny application | Satisfactory |
| Aritra Dutta | Apriori algorithm and shiny application | Satisfactory |
| Garima Singh | Apriori algorithm and shiny application | Satisfactory |
| Yangchen Sherpa | App deployment to cloud platform | Satisfactory |
| Kiran Negi | App deployment to cloud platform | Satisfactory |

**REFERENCES**

1. Joshi, S.M., 2018. Market-basket analysis using Apriori algorithm in data mining. *International Research Journal of Engineering and Technology*, *5*(4), pp.2861-2863.
2. Gupta, S. and Mamtora, R., 2014. A survey on association rule mining in market basket analysis. *International Journal of Information and Computation Technology*, *4*(4), pp.409-414.
3. KARAHAN ADALI, G. and BALABAN, M.E., 2019. A Dynamic Application of Market Basket Analysis with R and Shiny in The Electric Materials Sector. *International Journal of InformaticsTechnologies*, *12*(2).