

HEURISTIC APPROACH IN RETAIL SECTOR BY USING MARKET BASKET ANALYSIS TECHNIQUE

Abhang Mali^{*1}, Amol Dhawale^{*2}, Sumit Rokade^{*3}, Shraddha Ashtekar^{*4},
Prof. R. H. Borhade^{*5}

¹⁻⁴Students, Computer Engineering

⁵Head of Department of Computer Engineering Smt. Kashibai Navale College of Engineering
Pune, Maharashtra, India

Abstract - A company or organization analyzes their products to increase the product's deployment substantially. They also dive deep into knowing customer behaviour. In the same way, Market Basket Analysis involves the analysis of products to extract their relationship with each other. This aids the retailers to know which items or itemsets would be more frequently purchased. The paper includes the implementation of Market Basket Analysis technique on various grocery datasets. It includes different methodologies/tools used, Market Basket Analysis technique system architecture, data flow diagrams, and test cases. It distinctly explains the working of the Apriori algorithm. The deployment model here, represents working of the dynamic web application. It signifies how the Apriori algorithm helps to predict frequent itemsets. The performance metrics' outcomes exhibit the assurance of the model to be used in grocery shops to predict frequently purchased items.

Key Words: Association Rule Mining (ARM), Data Mining, Frequently purchased items, FP Growth Tree algorithm, Market Basket Analysis, Shopkeepers

1. INTRODUCTION

The retailers of small fields like general stores, medical stores, and grocery fields store their data either in a book(record) or in an Excel sheet. Here they invest their time and memory to analyse their present items to extract some useful insight. This insight is regarding the increase of particular products/items' sales.

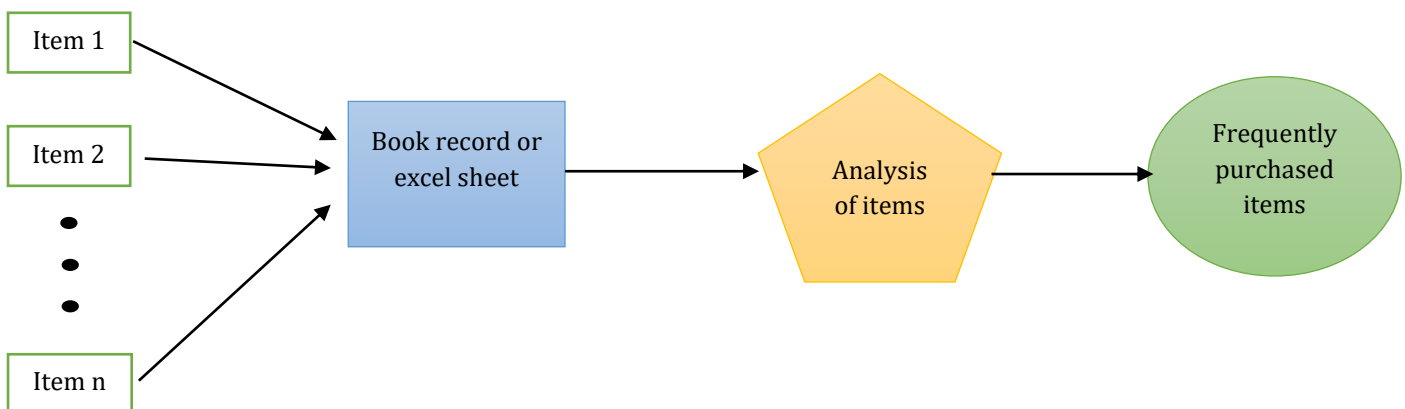


Fig1. Prediction of items without Market Basket Analysis

For example – a customer goes to a grocery shop. If he/she buys bread, then it is more favorable that he/she will also buy milk or butter with it. The shopkeeper keeps these items (milk, butter, jam, etc) close to the milk item. This works on the probability of these secondary items that are to be purchased along with the primary item. Thus, finding the probability and displaying frequently purchased secondary items is what Market Basket Analysis works on.

2. OBJECTIVE

To create a dynamic web application that predicts frequently purchased items with the help of the Market Basket Analysis technique.

3. LITERATURE SURVEY

References	Description	Gap Found
1	The Market Basket Analysis concept is explained at the foundational level.	Absence of brief explanation of the concept.
2	Basic working of Apriori and FP Growth algorithm is discussed.	Absence of other algorithms
3.	Apriori algorithm is explained and collaborative filtering concept is discussed in order to create a recommender system.	Drawbacks of Apriori algorithm are not discussed.
4.	Describes market share between different brands of sanitary napkins in a certain supermarket chain.	Lack of usage at large scale.
5.	Two algorithms namely Apriori and FP Growth are exclusively explained.	Applications of these algorithms are discussed.
6.	Terminologies and working of the Apriori algorithm are clearly discussed.	The drawbacks of Apriori algorithm are not discussed.
7.	An exclusive correlation is developed between data, information support, and confidence.	Absence of implementation of this concept at large scale.
8.	MBA technique uses innovative technologies to help retailers maintain their customer retention and satisfaction.	Since the Apriori algorithm has some drawbacks, it prevents the technique from being fully acceptable.
9.	New market basket analysis technique i.e., Logical Rule Mining(LRM) is explained here.	Since it is not in use at large, requires more awareness.
10.	A new method known artificial neural network technique is implemented in place of Apriori	It has more complexity as compared to Apriori algorithm.
11.	One of the applications of Market Basket analysis i.e., Image Based Product Recommendation is discussed here.	Faces overfitting issue.
12.	Market Basket Analysis is done on water resources in order to predict frequently occurring issues.	It is a time-consuming process.
13.	Market Basket Analysis technique is used in healthcare. the technique is used to find frequently occurring diseases.	Since, Apriori algorithm is used here, it leads to some drawbacks.
14.	Market Basket Analysis concept in sports field is thoroughly discussed here.	Algorithm used is not good enough for large datasets.
15.	Data mining and Market Basket Analysis are used to recognise patterns and then predict the outcome in order to increase sales of paper industry.	Need of an efficient algorithm for implementation like FP Growth Tree

The paper includes basic introduction of Market Basket Analysis technique. They also include basic terminologies of the technique, need of the technique. At last, applications of Market Basket Analysis are discussed. These include -

- Market Basket Analysis in Image Based Product Recommendation
- Market Basket Analysis in water resources.
- Market Basket Analysis in healthcare
- Market Basket Analysis in sports field.

4. METHODOLOGY

Market Basket Analysis is a method that helps retailers/organizations increase product sales/deployment. It performs two important functions which are as follows -

1. Finds the frequent items from the given transactions
2. Find relationships among various frequently occurring items.

The Market Basket Analysis is based on one important concept called Association Rule Mining (ARM). The ARM is an unsupervised technique. In other words, unlabelled data is provided as input to the model and then the model learns from this data to find the required useful insights. The association rule mining technique finds the hidden relationship among the products/items.

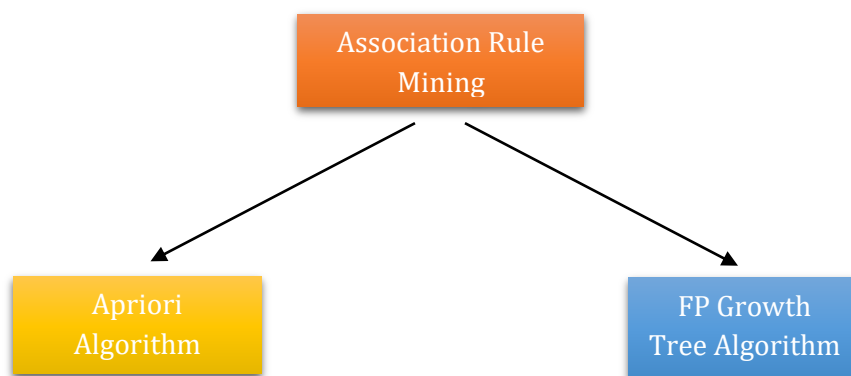


Fig 2. Algorithms of Association Rule Mining

4.1 Terminologies of Association Rule Mining(ARM) -

1. Support

It determines the number of occurrences of an item in overall transactions.

Support is calculated by this formula =
$$\frac{\text{Number of transactions containing the item}}{\text{Total number of transactions}}$$

2. Confidence

It determines the probability of purchasing the primary item along with its related item. Consider two items Milk and Bread. Here A is the primary item which is related to item B.

Confidence is calculated by this formula =
$$\frac{\text{Number of transactions containing both items Milk and Bread}}{\text{Number of transactions containing item Milk}}$$

3. Lift

It determines the probability of purchasing bread along with Milk.

Confidence is calculated by this formula =
$$\frac{\text{Number of transactions containing both items Milk and Bread}}{\text{Number of transactions containing item Bread}}$$

5. ALGORITHMS

5.1 Apriori Algorithm

The Apriori algorithm is a straightforward process. This means it is easy to understand the working of this algorithm. But the Apriori algorithm has some drawbacks –

1. Cannot or takes a considerable amount of time to process large transactions.
2. Faces computational complexity problems in handling large datasets.
3. Slow process

5.2 FP Growth Tree Algorithm

FP (Frequent Pattern) Growth algorithm is used in this project to implement the Market Basket Analysis technique. The FP Growth Tree algorithm overcomes all the drawbacks of the Apriori algorithm. The time complexity of FP growth algorithm is $O(n)$. This depicts that it has linear time complexity. In contrast to Apriori algorithm, FP growth algorithm works in the following steps: -

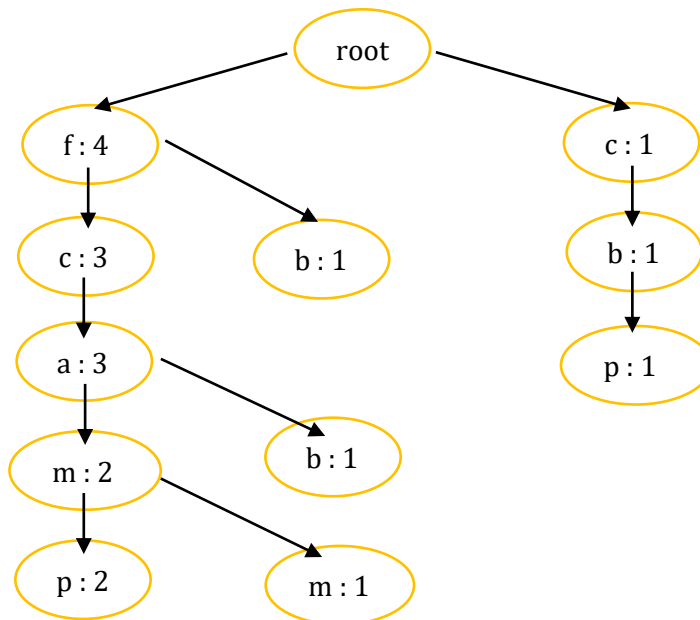


Fig 3. FP Growth Tree example

Step 1 -> Calculate the support counts of all items present in the given transactions.

Step 2 -> It creates a table that stores the support count of these frequent items in ascending order.

Step 3-> A Separate attribute/column is created for the conditional pattern base.

Step 4-> With the help of conditional pattern base data, a frequent pattern tree is generated.

Step 5-> Thus this tree will represent the frequently purchased items.

6. WORKING

In fig 4, there are multiple customers in one shopkeeper. A login system is provided for each and every shopkeeper so that they can accomplish their data. All data are stored inside the database so there is no need pen and paper. It decreases the time between items stored in the shop and the shopkeeper.

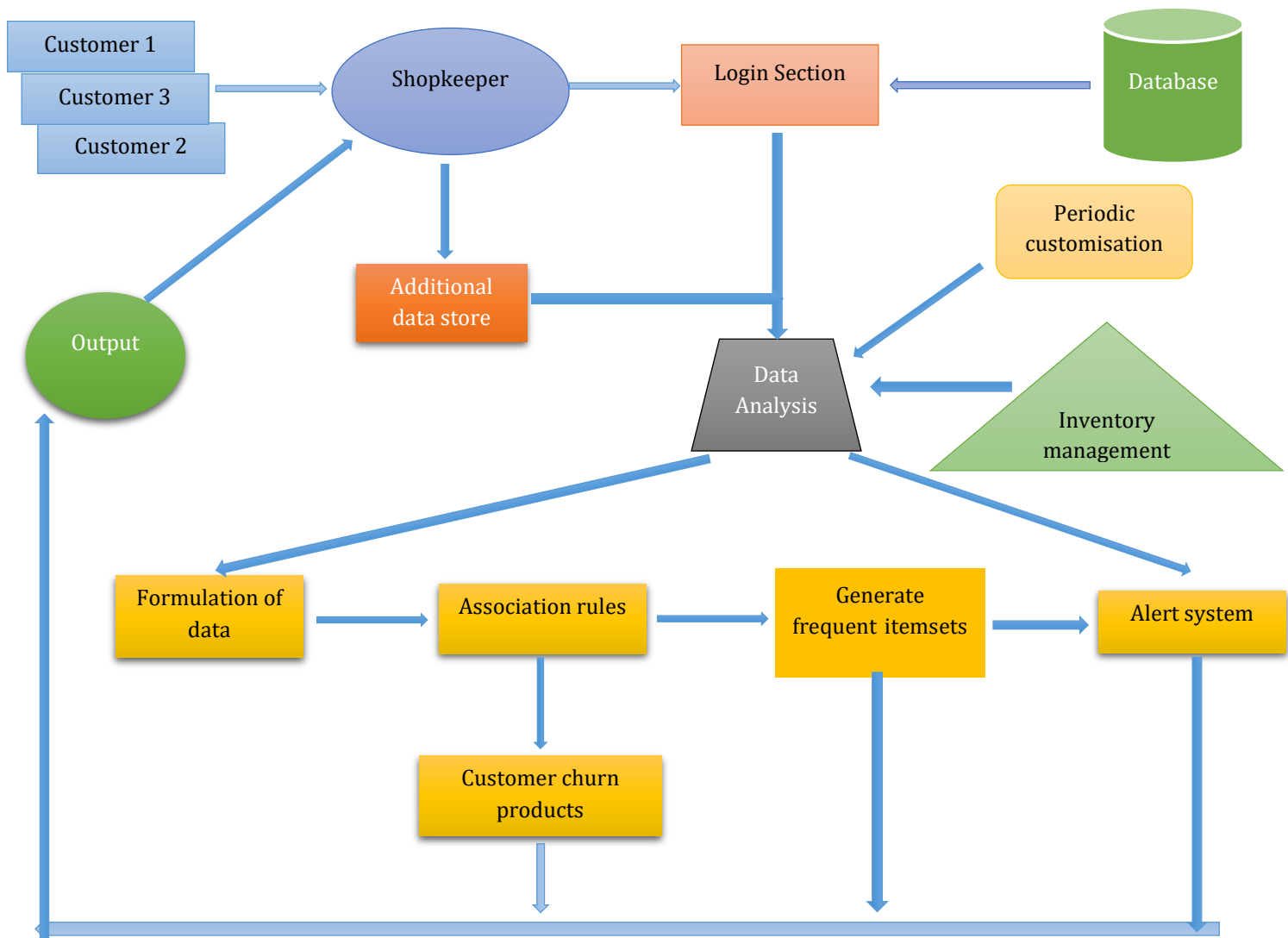


Fig 4. System Architecture of Market Basket Analysis Web App

1. Database

The database provides the facility of saving, deleting, updating, and reading the items from different customers. It also helps to organize the items in the shop which helps to increase the sales of shopkeepers.

2. Periodic Customization

Periodic customization is defined as a period between some months (one to four) in which the web application is again monitored and customized.

3. Formulation of Data

Formulation means a collection of data that can be achieved by the information of the customer purchased from the shopkeeper. Data contains the information of the customer along with items purchased by him or her and also it contains the information that is frequently purchased.

4. Association Rules:

This concept helps to generate the frequent items purchased by the customer. Information Gain gives the purity of an attribute that helps to understand the item's pattern. Entropy helps to understand the randomness or impurities of an attribute.

5. Customer Churn:

Customer churn means it is a collection of items that are not purchased by customers on certain period of time. It helps to remove unnecessary items from the shop and required items are stored in the shop. Customer churn can help to understand the customer and its purchasing pattern.

6. Alert System

It helps the shopkeeper understand how much stock remaining in his/her shop and understand. The user requirement according to customer to customer. Alert system provides the facility that which items are stored in the shop and which item removes from the shop.

7. Output:

Output report contains the items/itemsets that are frequently purchased by the customer in addition. The itemsets are processed by using the FP Growth Tree algorithm. Understanding the customer better helps customer behavior.

7. DATA FLOW DIAGRAM

Firstly shopkeeper needs to sign up by creating the username and password. These credentials are stored in the database. If shopkeepers login with authorized credentials then they can access to the services available to them. It helps to avoid the attack from different unauthorized user in our system. After successful login into the system, app will provide the services to the end user so that they can use this services and accomplished the result of it.

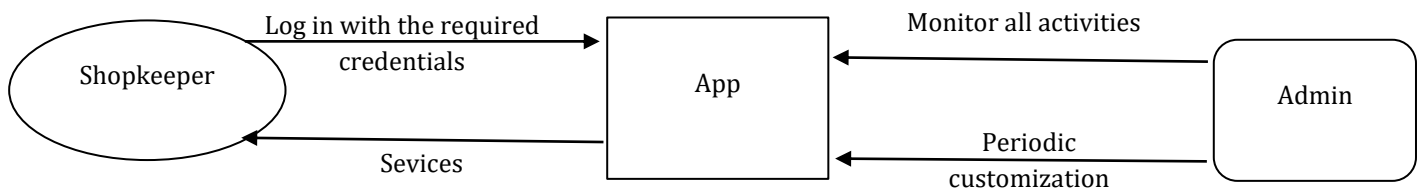


Fig 5. DFD Level 0

All app activities are monitored by the admin. Admin provides the support to the app so that they can handle the query of shopkeeper. Admin also helps in periodic customization of the app. In every date admin available for solving query of the shopkeeper and provide support to it.

In the fig6, data generation takes places by shopkeeper and customer purchasing items. Credentials contain the shopkeeper's name, password, username, email address and mobile number. After successful creation of the account of shopkeeper the app needs to fill the data about product present in the store. This data is stored inside the database using different methods like MySQL or PL SQL. Then from the database, data is sent for analysis and operation block. These data is observed by the admin and monitored by the shopkeeper.

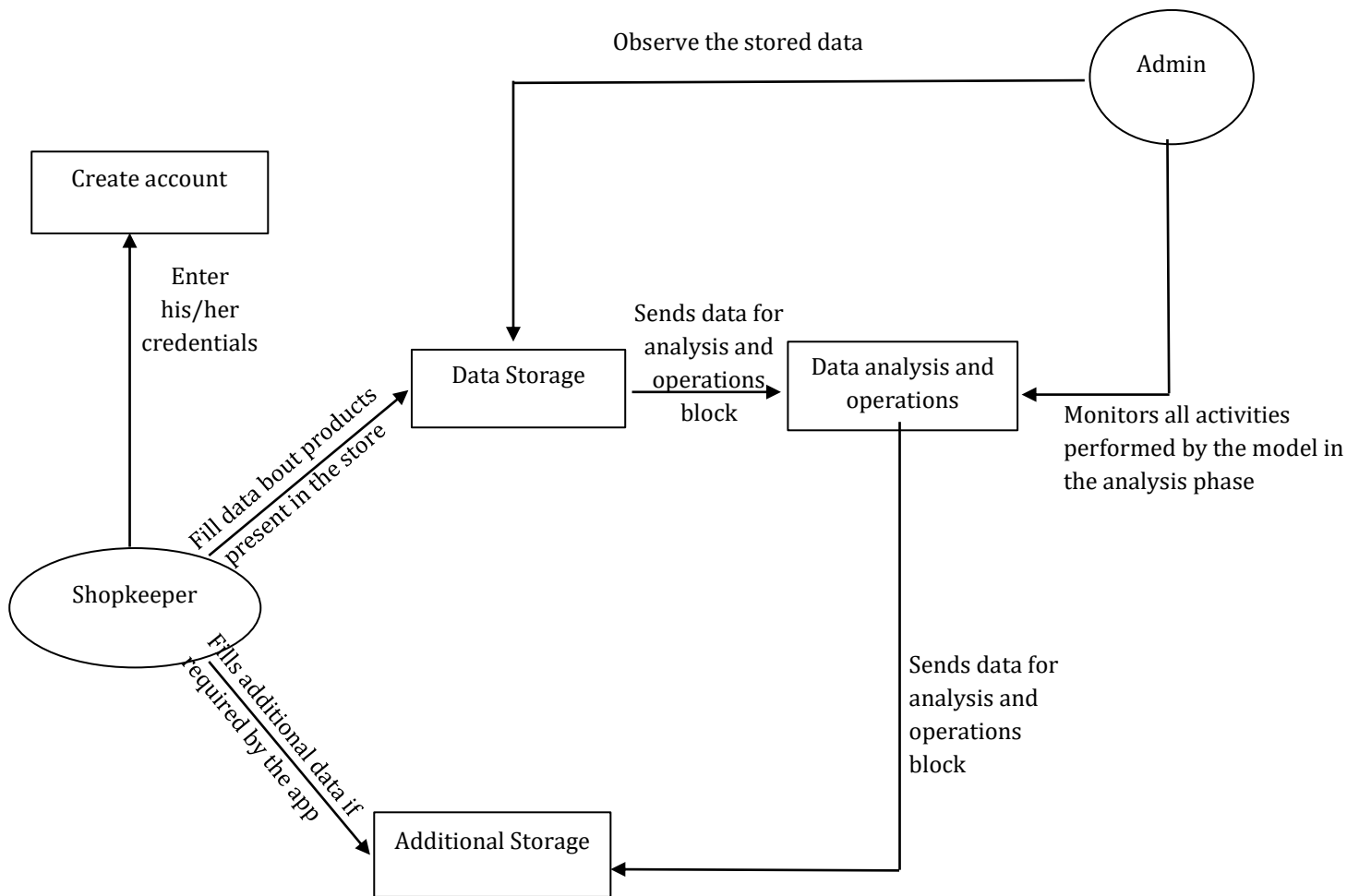


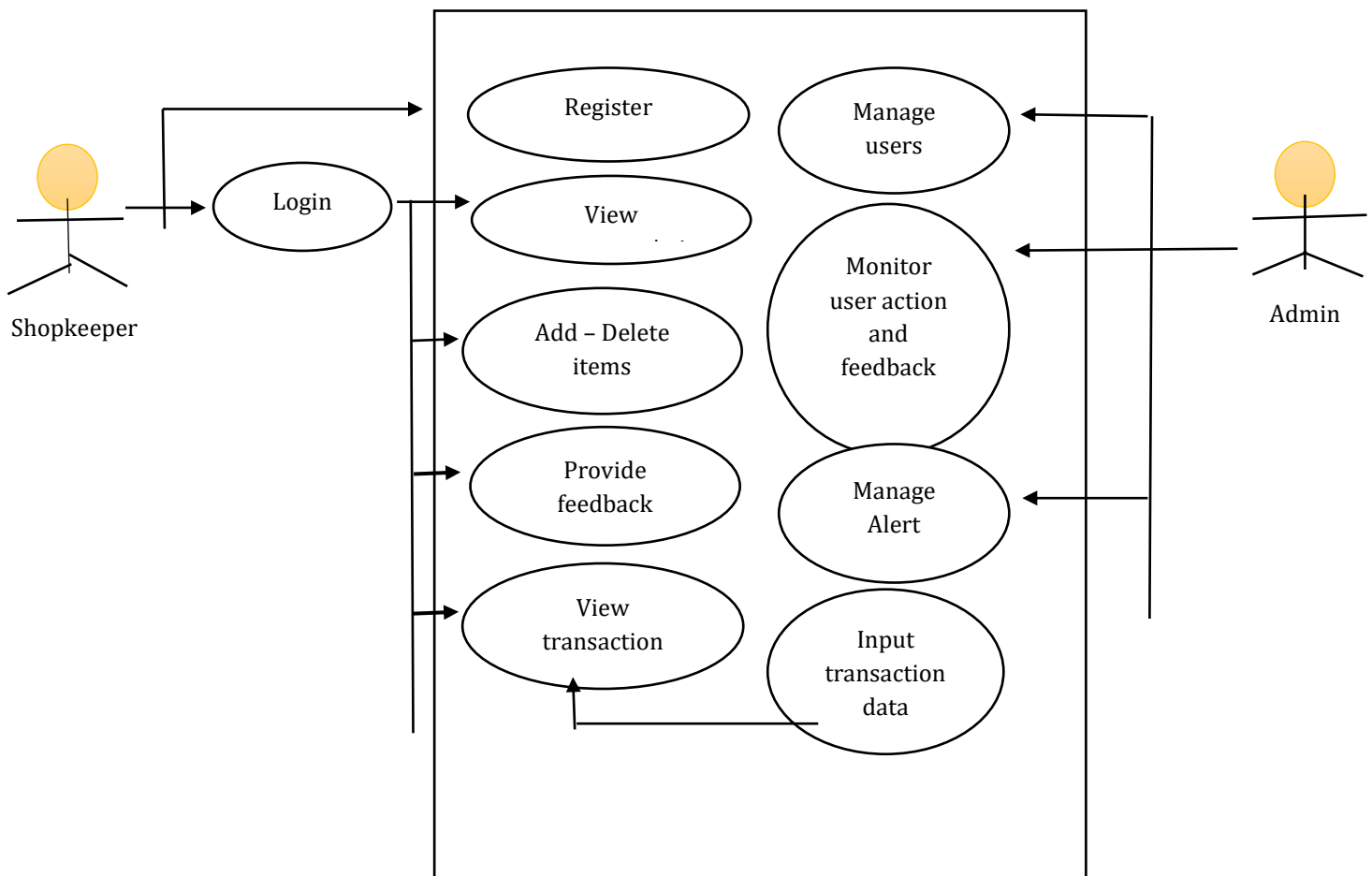
Fig 6. DFD Level 1

The data-related queries and defects are sent to the admin so that they can be able to fix it. Additional storage contains the information of the phone number all customer IDs etc. In the analysis phase, data is organized and monitoring is important; this can be done by the admin. The final report of the data is sent back to the shopkeeper so that they can analyze the report and perform certain actions on it. Action performance depends upon the working capacity of the data.

8. UML DIAGRAMS

Fig 7, the system has two actors: one is the shopkeeper and another is the admin. These two actors are responsible for data generation and performing or using the application. To use the application, firstly, the shopkeeper and admin need to login into the system by using credentials that include username, password, and email address. There are multiple shopkeepers but only one admin in the system.

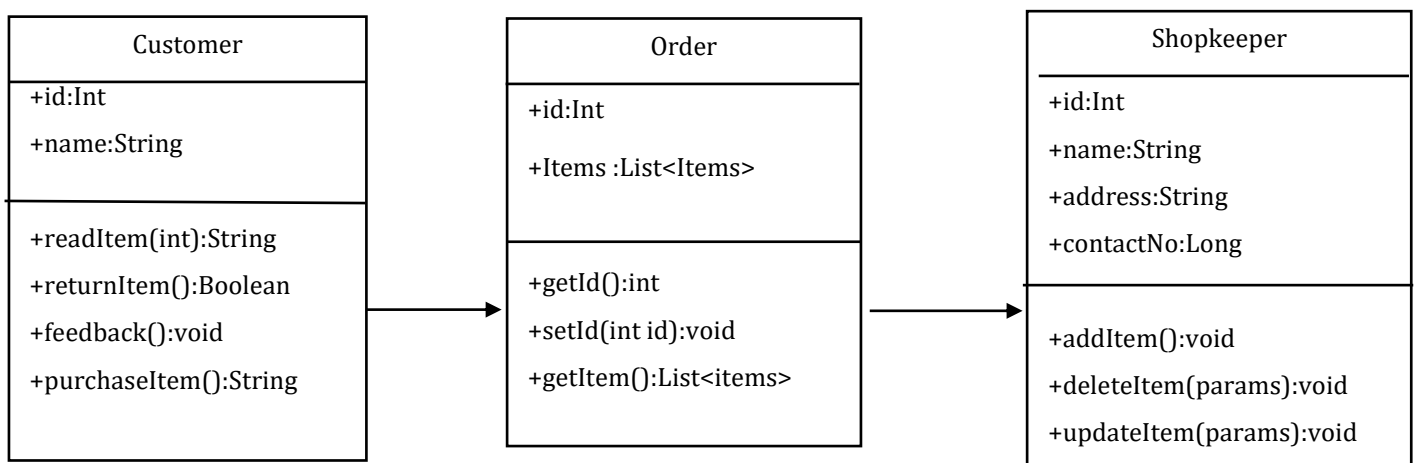
After successfully logging into the system, the shopkeeper can perform different operations like add, delete, update, and customer information like items purchased by the customer. To generate data, the shopkeeper is responsible.



Shopkeeper can view the recommendation provided by the Apps and along with perform the certain operation on the data. It have the facility to provide the feedback about the working of apps according to feedback provided by the shopkeeper. The user action perform by shopkeeper that manage and control by admin.

8.1 Class Diagram

Class Diagram gives the relationship between the different actors in the system. In addition it provide the facility of the understanding the attributes and method of the specific actors or class. It provides the facility of the abstraction and encapsulation so that the data is secure and robust. It cannot accessed by unauthorized person. Fig7. has different classes like customer, order, shopkeeper, admin and item etc.



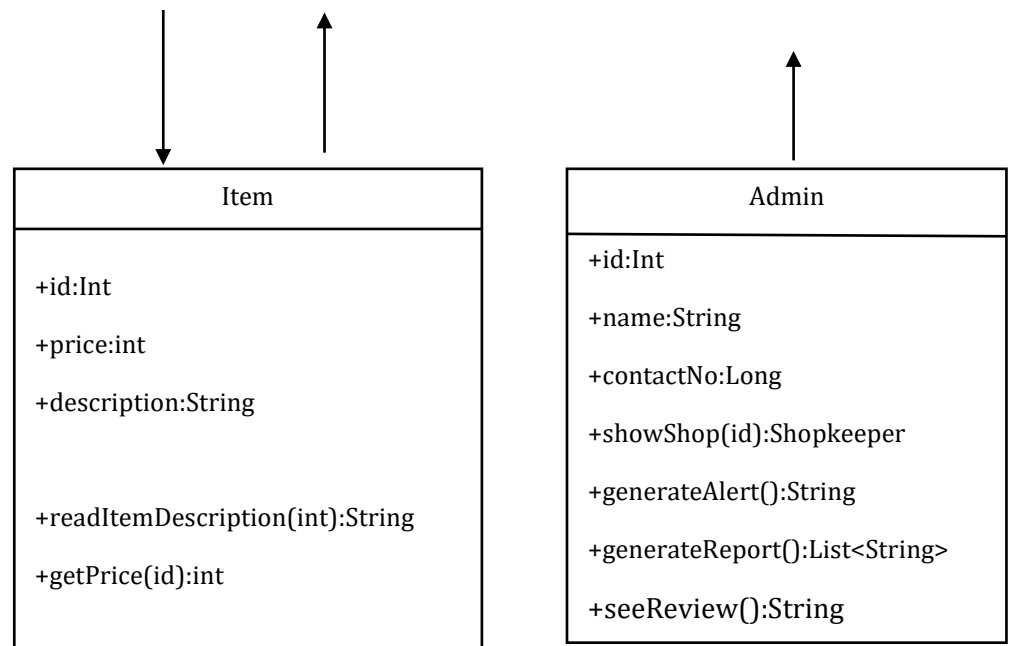


Fig 7. Class Diagram of Market Basket Analysis Webapp

Customer is responsible for the generation of the data. Customer has his/her id and name as an attribute. He/She can able to read the item in the store. If any item is not the favourite of any customer he or she can return the item back to the shopkeeper. The another classis order it contain basically two attribute id and items.

The items of the type list contain list of items that are purchased by the customer. He/She can able to read the item in the store. If any item is not the favourite of any customer he or she can return the item back to the shopkeeper. The another classis order it contain basically two attribute id and items. The items of the type list contain list of items that are purchased by the customer. Order class have the different method like setId and getId which is responsible for setting the id of the order and get the order id.

There is a main class called as shopkeeper that have the attribute id, name, string, address and contact number. This attribute have its own setter and getter method. It have method like for adding the the item have addItem method which takes the order id as a parameter. Another important character is the admin. It is responsible for generating the feedback and generating the report of the system. It also has id, name and contact number as a attribute.

8.2 Sequence Diagram

The sequence diagram gives the flow of the system how exactly works and which phase is going to be implemented in order to get the desired outcome. It shows the sequence of implementation of the data. In sequence diagram contains the information of different phases like actors, flow of data, application server, etc.

In fig 8, there are different actors like shopkeeper and different customer. Customer is responsible for generation of the data. In the shop multiple customers purchase the different types of items.

Shopkeeper requests for new login credentials. Shopkeeper performs the certain operation on the data like add the transaction data that contain the itemsets that are purchased by different customer. In addition of the analysis of data is done by the admin that operate and gives the final report to the application that is shopkeeper.

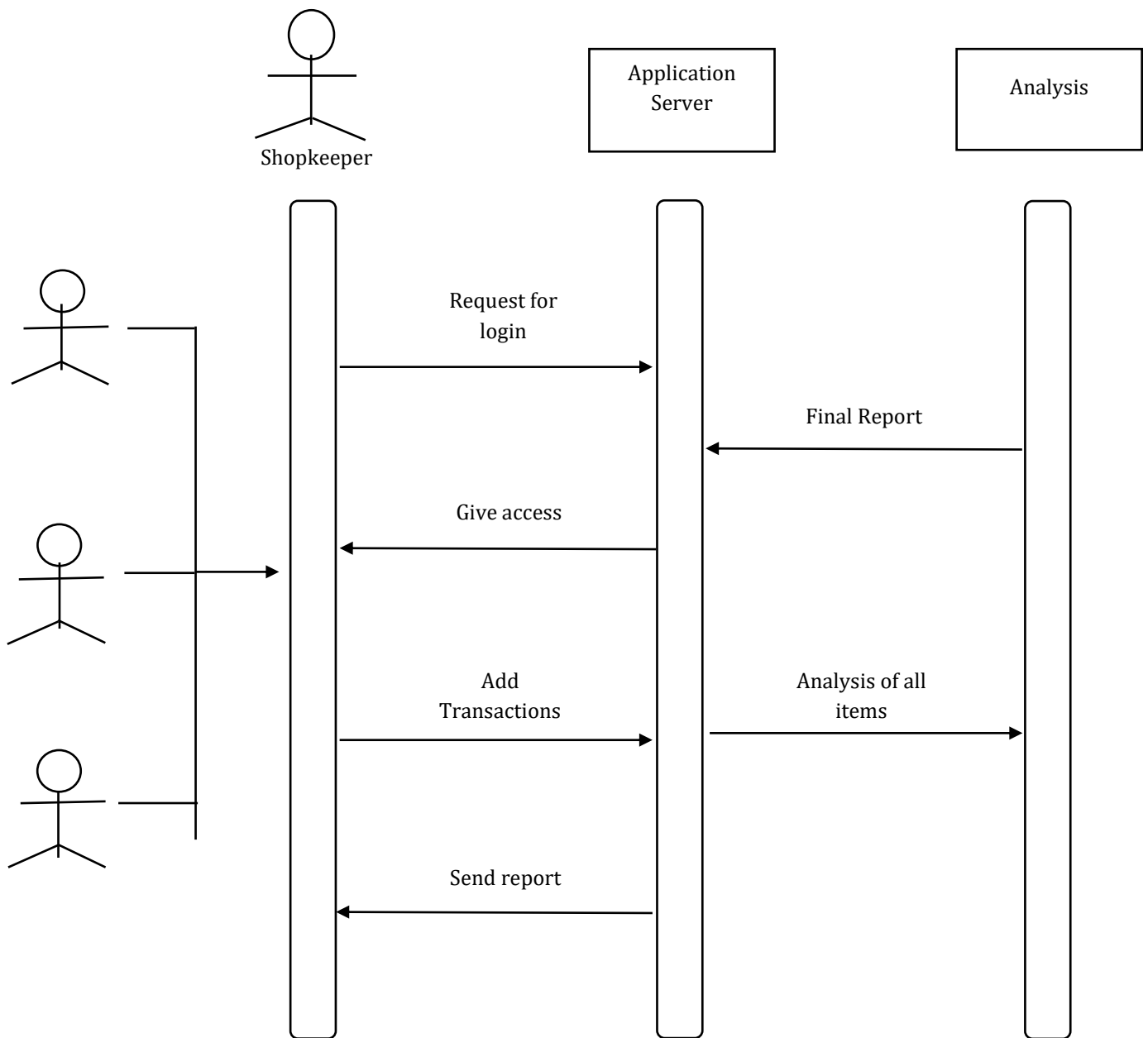


Fig 8. Sequence UML Diagram

9. DEPLOYMENT MODEL

All the data of user and the respective products/items is stored in cloud database. For this Firebase platform is used. Code of storing the data includes Firebase libraries. These are properly imported along with the initialize key of the web application. If any changes are required in the storing of the data like attributes names or number, then proper updation is done here.

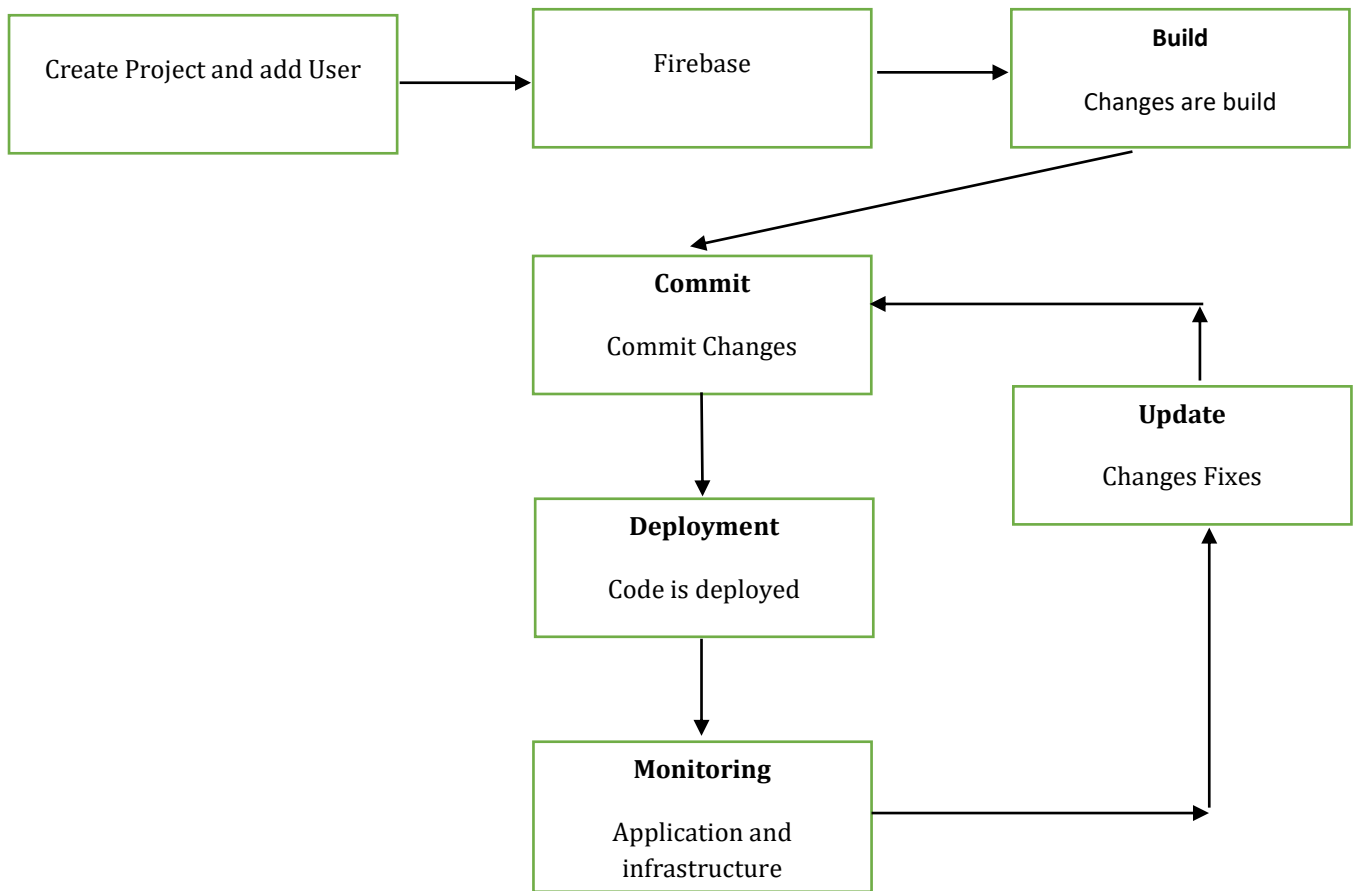


Fig 9. Deployment Model of Web App

The web application is first deployed in the staging environment where required customization or updation is performed. The results are then taken into consideration and then the web application is deployed.

10. RESULTS

Fig 10 represents a front view of the web application. This includes the signup and login options for the shopkeeper. The sidebar on the top left contains selection-boxes that include boxes as Storage, Account, Prediction, About, Sample respectively.

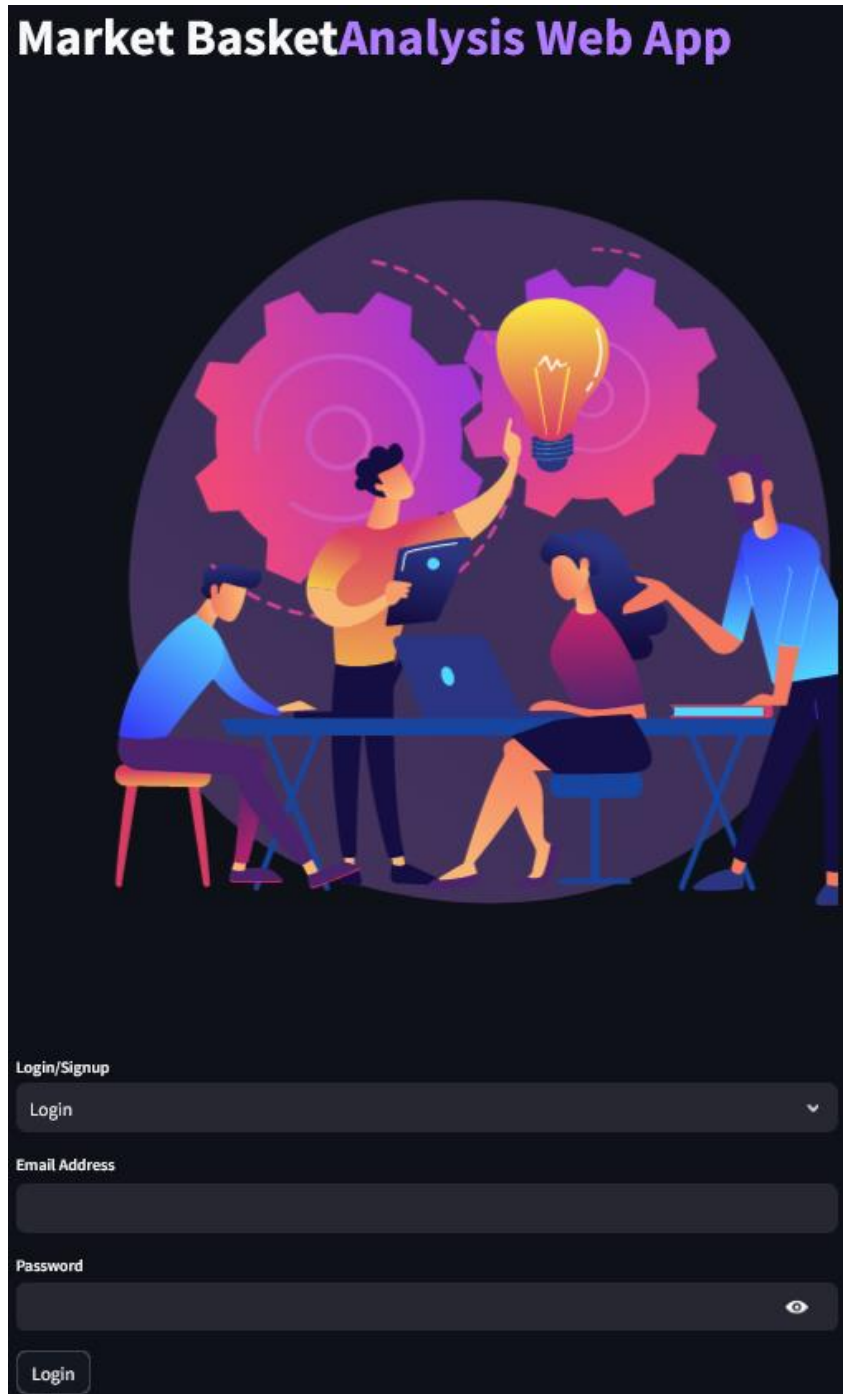


Fig 10. Account Page

Fig 11 represents the storage page. User when clicks on the “Storage” selection-box, then he/she will store their data here. The data can be in .csv, .exe, etc formats respectively. The data then will be stored in clouddata store of the Firebase platform.

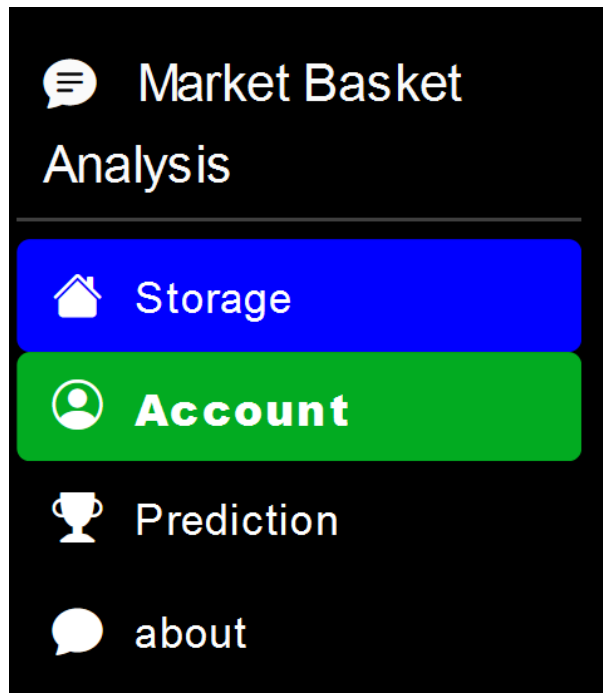


Fig 11. Storage page

Fig 12 represents the prediction page. After successfully storing the data, user will go on to the prediction page by clicking on the "Prediction" selection-box.

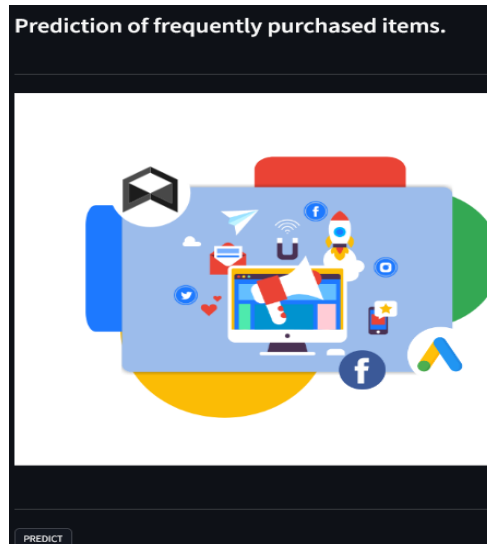


Fig 12. Prediction page

Fig 13 represents a dummy dataset being stored in the Firebase and then displayed on the prediction page.

Your dataset

	Member_number	Date	itemDescription
0	1,808	21-07-2015	tropical fruit
1	2,552	05-01-2015	whole milk
2	2,300	19-09-2015	pip fruit
3	1,187	12-12-2015	other vegetables
4	3,037	01-02-2015	whole milk
5	4,941	14-02-2015	rolls/buns
6	4,501	08-05-2015	other vegetables
7	3,803	23-12-2015	pot plants
8	2,762	20-03-2015	whole milk
9	4,119	12-02-2015	tropical fruit

Fig 13. Prediction output

Fig 14 represents the graph. The graph shows frequently purchased items. Item names are displayed below each bar of the graph. The size of the graph can be increased to represent more number of frequently purchased items. With the help of this graph, user (shopkeeper) will easily.

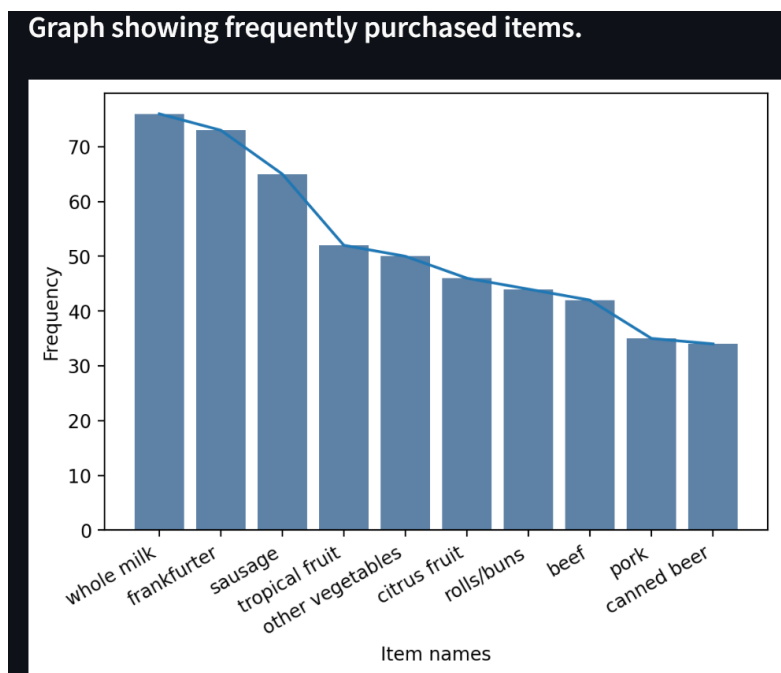


Fig 14. Graph showing frequently purchased items

11. CONCLUSION

The research involves deep analysis of Market Basket Analysis and the developed web application. Came to know different reviews of researchers on the Market Basket Analysis technique. Market Basket Analysis (MBA) is a powerful data analysis method with several significant benefits and drawbacks. Successfully created a dynamic web application called “Market Basket Analysis WebApp” that accepts customers’ data and efficiently predicts frequently purchased items. By analyzing deep into the concept, many new aspects of the technique were discovered. Also came to know the problems of the Apriori algorithm that were solved by FP Growth Tree algorithm. When applied correctly and in combination with other data

analysis techniques, market basket analysis is a useful technique. Thus, the Market Basket Analysis technique helps to increase the deployment rate of products at a substantial rate.

12. REFERENCES

- [1] Abhang Mali, Amol Dhawale, Sumit Rokade, Shraddha Ashtekar, Prof. R. H. Borhade, "HEURISTIC APPROACH IN RETAIL SECTOR BY USING MARKET BASKET ANALYSIS TECHNIQUE", International Research Journal of Modernization in Engineering Technology and Science, Vol 5, No. 11, 1615-1619. 2023.
- [2] Vippala Nagendra Reddy, Panta Sai Sathvik Reddy, "Market Basket Analysis Using Machine Learning Algorithms", 2021.
- [3] Yüksel Akay Ünvan, "Market basket analysis with association rules", 2020.
- [4] Shish Kumar Dubey, Sonu Mittal, Seema Chattani, Vinod Kumar Shukla, "Comparative Analysis of Market Basket Analysis through Data Mining Techniques", 2021.
- [5] Jiangping Chen, Xiaoxian Yang, Lihua Chen, Li Dong, Yuanyan Fu, "An Analysis of Shopping Basket in a Supermarket", 2010.
- [6] Maliha Hossain, A H M Sarowar Sattar, Mahit Kumar Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm", 2019.
- [7] Warnia Nengsih, "A Comparative Study on Market Basket Analysis and Apriori Association Technique", 2015.
- [8] Djoni Haryadi Setiabudi, Gregorius Satia Budhi, Mining Market Basket Analysis' Using Hybrid-Dimension Association Rules, Case Study in Minimarket X", 2011.
- [9] Ahmad M. A. Zamila, Ahmad Al Adwanb, T. G. Vasista, "Enhancing Customer Loyalty with Market Basket Analysis Using Innovative Methods: "A Python Implementation Approach", 2020.
- [10] Damla Oguz, Fatih Soygazi, "An interestingness measure for knowledge bases", 2023.
- [11] Anshul Bhargav, Robin Prakash Mathur, Munish Bhargav, "Market Basket Analysis using Artificial Neural Network", 2014.
- [12] Premanand Ghadekar, Anay Dombe, "Image-Based Product Recommendations Using Market Basket Analysis", 2019.
- [13] Tiyyasha Tiyyasha, Suraj Kumar Bhagat, Firaol Fituma, Tran Minh Tung, Shamsuddin Shahid, And Zaher Mundher Yaseen, "Dual Water Choices: The Assessment of the Influential Factors on Water Sources Choices Using Unsupervised Machine Learning Market Basket Analysis", 2021.
- [14] Abishek B. Rao, Jammula Surya Kiran, Poornalatha G, "Application of market-basket analysis on healthcare", 2021.
- [15] Wan Faezah Abbas, Nor Diana Ahmad, Nurlina Binti Zaini, "Discovering Purchasing Pattern of Sports Items Using Market Basket Analysis", 201
- [16] M. J. Zaki and W. J. Meira, Data Mining and Machine Learning Fundamental Concepts and Algorithms, vol.53, no. 9. 2020.