

# Market basket insights

Abstract

with good accuracy is essential for efficient algorithm. The issues for a leading supermarket are addressed here using frequent item set mining.

The project uses file as database. Here, the itemsets and transactions of items are kept in a matrix form representing rows as list of items and column as transactions.

The frequent item sets are mined from database using the Apriori algorithm and then the association rules are generated.

The project is beneficial for supermarket managers to determine the relationship between the items

# Introduction

- Background :
- Market Basket analysis is a data mining method focusing on discovering purchase patterns of the customers by extracting association or co-occurrences from a store's transactional data. For example, when the person checkout items in a supermarket all the details about their purchase goes into the transaction database. Later, this huge data of many customers are analyzed to determine the purchasing pattern of customers. Also decisions like which item to stock more, cross selling, up selling, store shelf arrangement are determined
- Association rule mining (ARM) identifies the association or relationship between a large set of data items and forms the base for market basket analysis. Association rule mining has been widely used in various industries besides supermarkets, such as mail order, telemarketing production, fraud detection of credit card and e-commerce.
- One of the challenges for companies that have invested heavily in customer data collection is how to extract important information from their vast customer databases and product feature databases, in order to gain competitive advantage. Market basket analysis has been intensively used in many companies as a means to discover product associations.
- A retailer must know the needs of customers and adapt to them. Market basket analysis is One possible way to find out Which item can be put together
- Market Basket Analysis helps to identify the purchasing behavior of the customer. By mining the data from the huge transaction database shop managers can study the behavior or buying habits of the customer to increase the sale. In Market Basket Analysis, you look to see

# Problem statement

- Nowadays people buy daily goods from super market nearby. There are many
- supermarkets that provide goods to their customer. The problem many retailers face is the
- placement of the items. They are unaware of the purchasing habits of the customer so
- they don't know which items should be placed together in their store. With the help of
- this application shop managers can determine the strong relationships between the items
- which ultimately helps them to put products that co-occur together close to one another.
- Also decisions like which item to stock more, cross selling, up selling, store shelf arrangement are determined

# Objectives

- . To identify the frequent items from the transaction on the basis of support and confidence
- b. To generate the association rule from the frequent item sets.

# Scope

- Scope of the application is limited to desktop application right now.  
The application is targeted towards a Supermarket to nepal

# Limitations

- A. This application will be desktop and will not be available online.
- b. Input to the application will be a file which contains integer values representing the list of items , the integer value Will be mapped Manually .

# Report organization

- Preliminary Section
- ▢ Title Page
- ▢ Abstract
- ▢ List of figures
- ▢ List of Tables
- 
- Introduction Section   ▢ Background
- ▢ Problem Statement
- ▢ Objectives
- ▢ Scope and Limitation
- 
- Literature Review Section   ▢ Apriori Algorithm
- ▢ FP-growth
- Methodology Section   ▢ Data collection
- ▢ Data pre-processing
- ▢ System design
- 
- Result Section   ▢ Test Cases
- ▢ Testing correctness of output
- 
- Conclusion Section   ▢ Conclusion
- ▢ Recommendations



# Data processing

- The data collected was mapped manually as integer values as shown in Figure 4. For
- example the “Fruit” was labeled as 1, “Bread” as 2 “Soups” as 4 and so on.
- 1,fruit
- 2, Bread
- 4, soups
- 6, yougurt
- The mapped integer’s values were then saved in a text file and given as the input to the
- system. Figure 5 shows the input file that is given to the system.

# Support

- The support of an item is the number of transaction containing the item. Those items that
- do not meet the minimum support are excluded from the further processing. Support
- determines how often a rule is applicable to a given data set.
- $\text{Support}(X \cup Y) = \min(\text{Support}(X), \text{Support}(Y))$

# Pseudo code

- /Find all frequent itemset
- Apriori(database D of transaction, min\_support){
- $F_1 = \{\text{frequent 1-itemset}\}$
- $K=2$
- While  $F_{k-1} \neq \text{Empty Set}$
- $C_k = \text{AprioriGeneration}(F_{k-1})$  //Generate candidate item sets.
- For each transaction in the database D {
- $C_t = \text{subset}(C_k, t)$
- For each candidate  $c$  in  $C_t$ {
-

This project is implemented in java. For the user interface to provide the input data java swing is used to design the interface. The apriori algorithm is used to process the data and generated the association rule as a output in a file.

- 4.1.1 Tools used

- IntelliJ: IntelliJ was used as a IDE to develop this application

- Java swing: Java swing was used for designing the user interface.

- Java: java programming language was used to implement the algorithm

- File: File is used as the database to process the data.

- 

- 4.1.2 Description of major classes and methods.

- 

- MainWindow

- This class is used to display the user interface for providing the input to the system. This

- class extends the JFrame class

- Some of the important methods of thActionPerformed()

- askUserToChooseInputFile()

- askUserToChooseOutputFile()

- 

- AlgoApriori

- This is the main class that executes the algorithm after user presses the run algorithm

- button.This class performs the processiong of the data from user input and display output

- to the user decided output file.

- Some of the important methods of this class are:

- runAlgorithm(double minsup, String input, String output)

- generateCandidate2(List<Integer> frequent1)

- generateCandidateSizeK(List<Items et> levelK\_1)

- printStats()

# Testing

- The purpose of testing is to discover errors. Testing is the process of trying to discover
- every conceivable fault or weakness in a work product. It provides a way to check the
- functionality of components, subassemblies, assemblies and/or a finished product. It is the
- process of exercising software with the intent of ensuring that the Software system meets
- its requirements and user expectations and does not fail in an unacceptable manner.
- Unit testing was performed to test correctness of different modules.

# Conclusion

- The Apriori algorithm effectively generates highly informative frequent itemsets and association rules for the data of the supermarket. The frequent data items are generated from the given input data and based on the frequent item sets strong association rules were generated.
- 
- Recommendations
- The input data given to the application is used as the integer value mapped from the transaction database. The mapping is done manually. If database converter is made then the system will work effectively for any format of data. The application can be efficiently used by using more efficient algorithm rather than Apriori Algorithm in future.

