Product Recommendation System

A

Project Report

submitted in partial fulfillment of the
requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE

Specialization in

Open Source And Open Standards

By :

Name	Roll No
Tushar Singhal	R100218063
Vidhu Sharma	R100218066
Vanshul Gupta	R100218071
Reet Dubey	R100218076

Under the guidance of

Mr. Arjun Arora

Assistant Professor
Department of Cybernetics

Department of Cybernetics
School of Computer Science
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Bidholi, Via Prem Nagar, Dehradun, Uttarakhand

CANDIDATES DECLARATION

I/We hereby certify that the project work entitled **Product Recommendation System** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science And Engineering with Specialization in Open Source And Open Standard and submitted to the Department of Cybernetics at School of Computer Science, University of Petroleum And Energy Studies, Dehradun, is an authentic record of my/ our work carried out during a period from **August, 2020** to **December, 2020** under the supervision of **Mr. Arjun Arora ,Assistant Professor ,Department of Cybernetics**

The matter presented in this project has not been submitted by me/ us for the award of any other degree of this or any other University.

Tushar Singhal (R100218063) **Vidhu Sharma(R100218066)**

Vanshul Gupta(R100218071) **Reet Dubey (R100218076)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Date: 05 December 2020)

(Mr. Arjun Arora)
Project Guide

Dr. Monit Kapoor

Head
Department of Cybernetics
School of Computer Science
University of Petroleum And Energy Studies
Dehradun - 248 001 (Uttarakhand)

ACKNOWLEDGEMENT

We wish to express our deep gratitude to our guide **Mr. Arjun Arora** , for all advice, encouragement and constant support he has given us through out our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank to our Head of the Department, **Dr. Monit Kapoor** , for his great support in doing our **Product Recommendation System** at **SoCS**.

We are also grateful to **Dr. Manish Prateek Professor and Dean CoES**, UPES for giving us the necessary facilities to carry out our project work successfully.

We would like to thank all our **friends** for their help and constructive criticism during our project work. Finally we have no words to express our sincere gratitude to our **parents** who have shown us this world and for every support they have given us.

Name	**Tushar Singhal**	**Vidhu Sharma**	**Vanshul Gupta**	**Reet Dubey**
Roll No.	**R100218063**	**R100218066**	**R100218071**	**R100218076**

ABSTRACT

Implementation of the Product Recommendation System which automatically recommends the products related to the searched or frequently bought products. For the implementation we will use Apriori Algorithm. The Apriori Algorithm is an influential algorithm designed to operate on database containing business transactions for analysing frequent itemsets and extending them to larger item-sets as long as those item sets appear frequently often in that database, individually. It uses bottom up approach where frequent subsets are extended one item at a time.
Using this algorithm, the number of item sets that has to be examined can be sorted and the list of popular item-sets can be obtained. Hence, Apriori can be applied to find frequent item-sets season-wise. Further, the efficiency of the algorithm can be improved using the following variations

- Hash based technique- a K- item set whose corresponding hashing bucket count is below the threshold cannot be frequent.
- Partitioning-Any item set that is potentially frequent in database must be frequent in at least one of the partitions of the database.
- Sampling-Mining on a subset of given data, lower support threshold + a method to determine the completeness.

Keywords: Frequently Bought, Bottom up Approach

TABLE OF CONTENTS

Contents

LIST OF FIGURES

List of Figures

1 Introduction

Our project is based on the product recommendation system. It can be used to recommend products to customers of a general store on the basis of the frequently bought items in a season. The project uses the Apriori algorithm which is used to calculate the items that are bought more than the stated support count. It also uses techniques like hashing and partitioning to increase its efficiency.

2 Literature Review

Apriori algorithm can be used in various fields including Market Basket Analysis. This process analyzes what items are frequently bought together by the customers. Using this, shop owners can also develop marketing strategies as to which product should be placed where in a supermarket. Association rules analysis is a technique to uncover how items are associated to each other. There are two common ways to measure association.
Measure 1: Support
Support count gives the indication of much an items is being bought. If an item is bought at an amount that increases our profit then it can be considered as a support threshold.

Measure 2: Confidence.
Confidence is an indication of how often the rule has been found to be true.
Rules that satisfy both a minimum support threshold (min sup) and a minimum confidence threshold (min conf) are called strong.
Apriori property: All non empty subsets of a frequent itemset must also be frequent.
It has 2 steps:
Step 1: Apply minimum support to find all the frequent sets with k items in a database.
Step 2: Use the self-join rule to find the frequent sets with k+1 items with the help of frequent k-item sets. Repeat this process from k=1 to the point until we are unable to apply the self-join rule. This approach is called the “bottom up” approach.

3 Problem Statement

Recommendation of frequently bought items in a general store over a period of time using Apriori algorithm and optimizing the efficiency by various techniques such as hashing, partitioning and sampling.

4 Objective

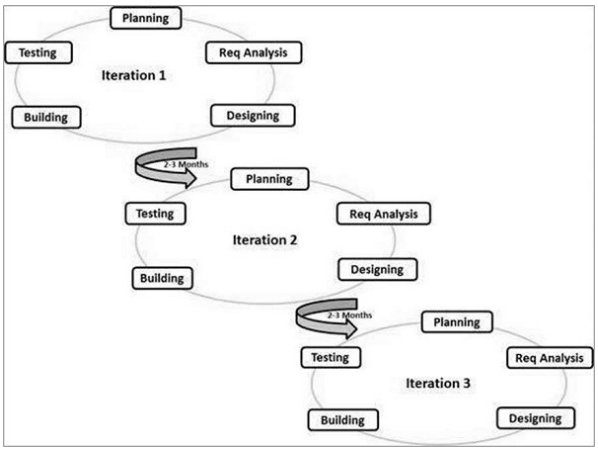
1. To recommend products to customers of a general store using frequent item-sets obtained season-wise.

2. Methods to improve Apriori’s efficiency:
    i. Hash based item set counting: A k –item set whose corresponding hashing bucket count is below the threshold cannot be frequent.
    ii. Partitioning: Any item set that is potentially frequent in database is must be frequent in at least one of the partitions of database.
    iii. Sampling: Mining on a subset of given data

5 Design Methodology

The methodology we are using for our project is Agile development. Agile is a process by which a team can manage a project by breaking it up into several stages and involving continuous improvement and iteration at every stage.

Agile model allows us to change the requirements after the development process starts, so it is suitable for our project. Agile SDLC model is a combination of iterative and incremental process models with focus on process adaptability and customer satisfaction by rapid delivery of working software product. Agile Methods break the product into small incremental builds. These builds are provided in iterations. Every iteration involves cross functional teams working simultaneously on various areas like – planning, Design. coding etc



Agile model

6 Implementation

The system is programmed in such a way to recommend the frequently bought item and increase the sale of store. This is achieved by using various approaches such as hashing to count the minimum support count so that the system compares it previous list and recommend the maximum selling product to the person.

6.1 Pseudocode

Apriori Algorithm

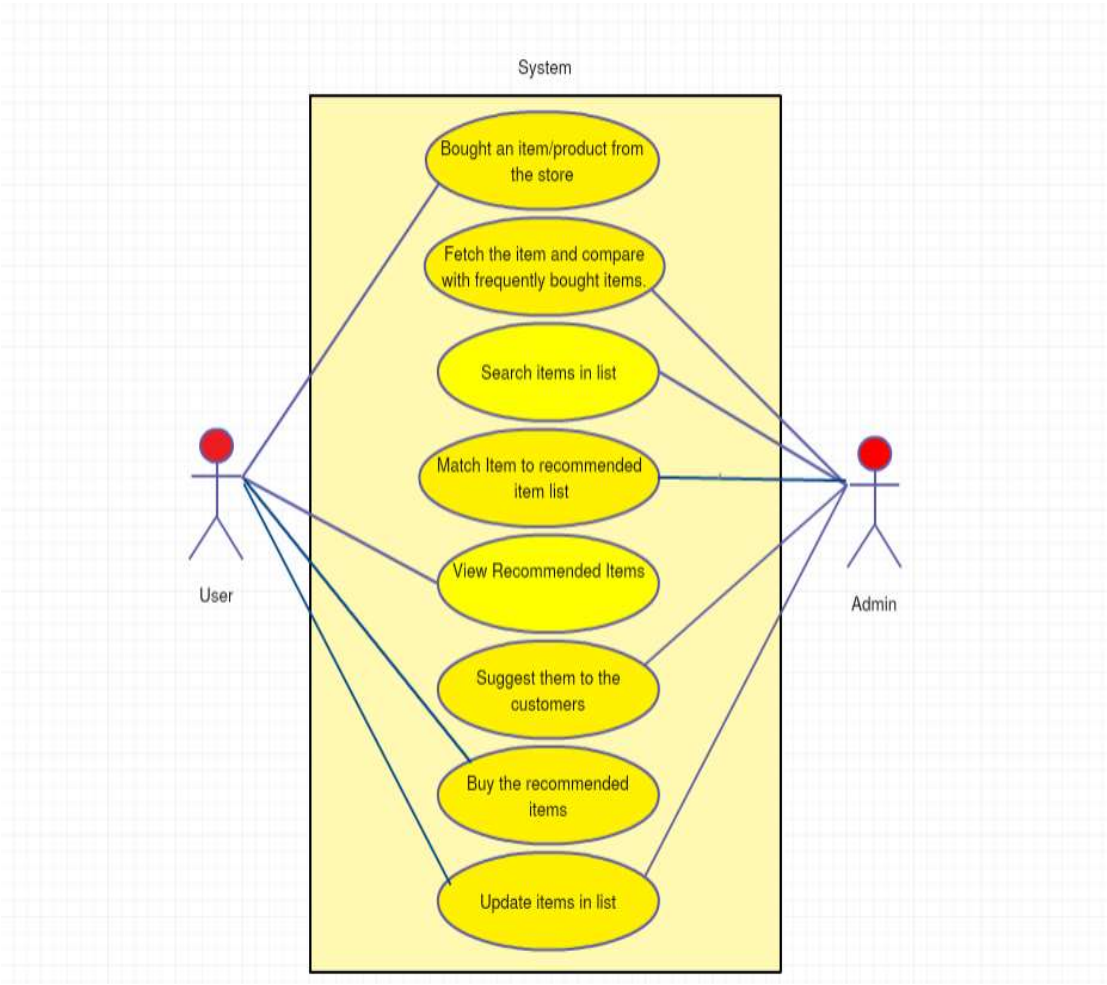
C_k : Candidate itemset of size k
L_k : Frequent itemset of size k

L_1 = {frequent items};
for(k=1; L_k ! = Φ ; k++) do begin
C_{k+1} = candidates generated from L_k;
for each transaction t in database do increment the count of all candidates in C_{k+1} that are contained in t
L_{k+1} = Candidates in C_k with min_support
end

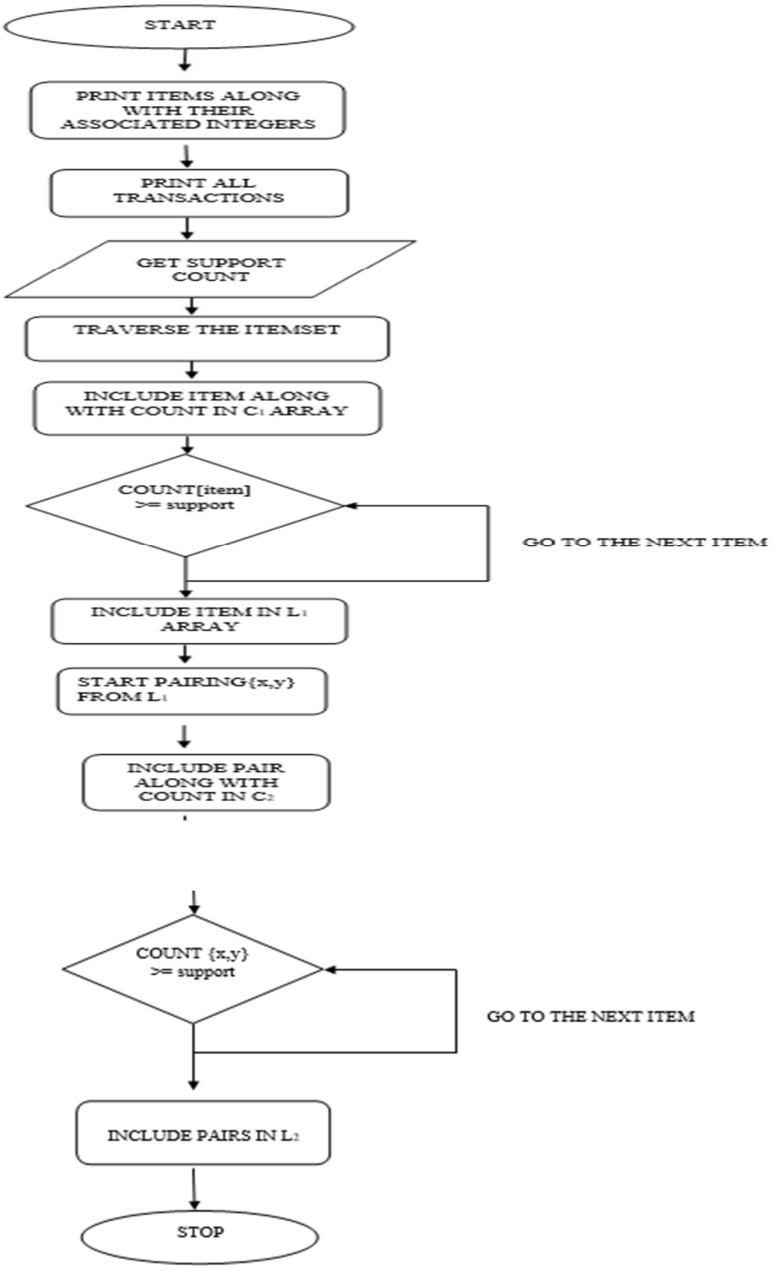
return U_k L_k...

6.2 Diagrams

Use case diagram

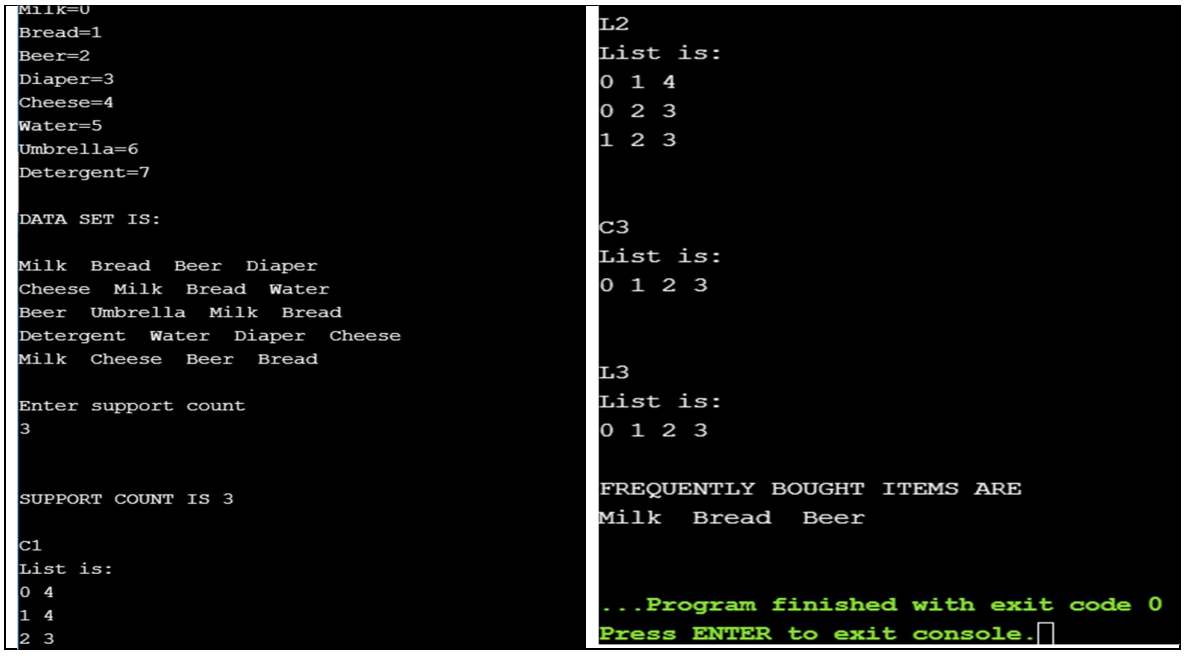


Activity diagram

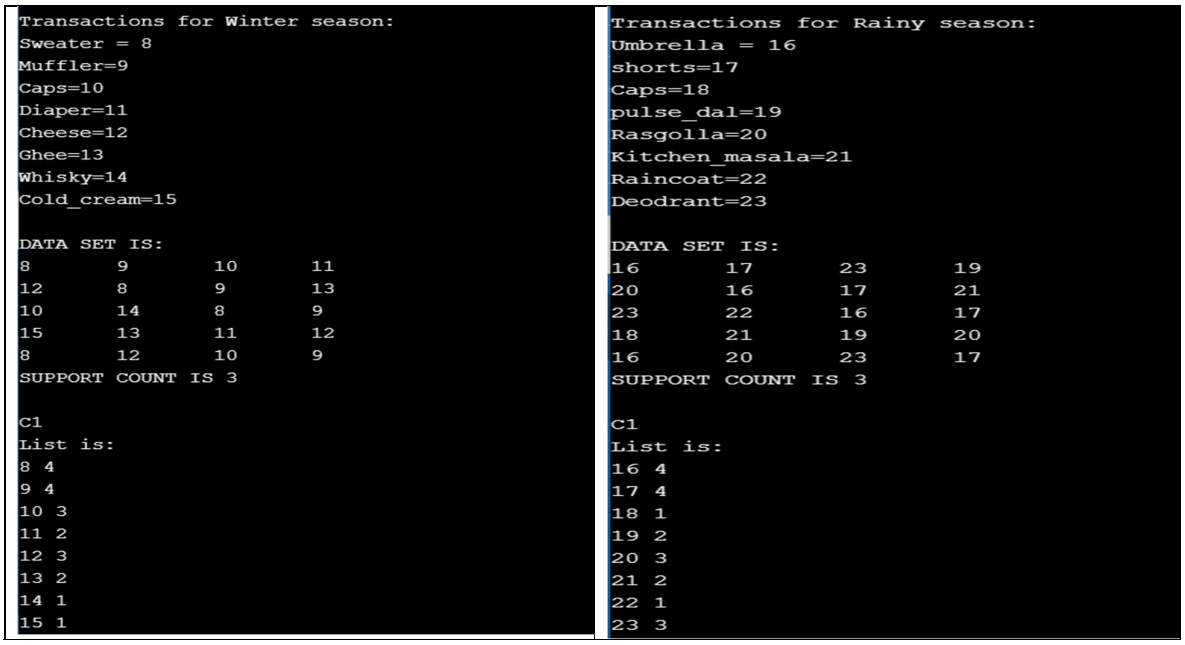


6.3 Output Screen

We have the basically two output ,one is based on person current items recommendation that what other people frequently buys with those item that you buy ,second is according to season that which item is frequently bought in this season example if a person want to buy bread and milk our system recommend him that butter is frequently bought with bread and milk and also if the season is winter then our system recommend him to buy honey and respectively other according to their purchase and seasons.



1 output figure (Item wise)



2 output figure (season wise)

6.4 Result Analysis

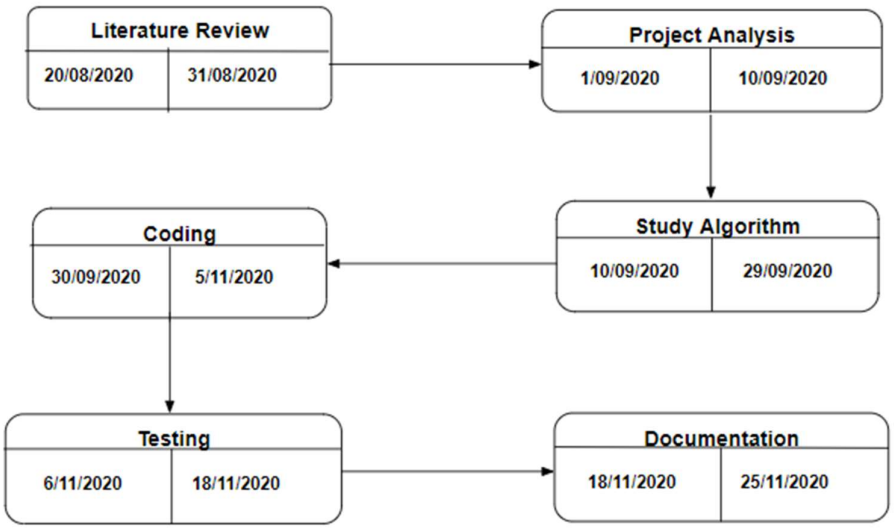
The result consist of Recommending the frequently bought item to buyer and also recommends the frequently bought item according to the season .We use apriori algorithm over any other algorithm as time complexity and space complexity of aprori in this situation is best .In this we have made our sample data set with the help of array to make the list of items separately.

7 Conclusion and Future Scope

The program is design in such a way that it recommends the frequently bought product to the person, we have used aprori algorithm and various other techniques such as hashing, sampling, partitioning to get our desire output. The future scope of this project is

- It provides customer the more relevant products
- Helps in boosting the sale of store owner
- Improves business strategies

8 Pert Chart



References

[1] Jiawei Han and Micheline Kamber, Data Mining: Concepts And Techniques, 2 nd ed. SanFrancisco (CA) : Diane Cerra, 2006. [Online] Available : Bharat University ebook.

[2] Zhihua Cai , Zhenhua Li, Zhuo Kang, "Computational Intelligence and Intelligent Systems" 4th International Symposium on Intelligence Computation and Applications, ISICA 2009, Huangshi, China pg no. 438-440 on Implentation Of Apriori Algorithm

[3] Jatinder Kaur, Rajeev Kumar Bedi, S.K. Gupta, 1192 International Journal of Computer Sciences and Engineering Vol -6, Issue -6 Review Paper on "Product Recommendation Systems a Comprehensive Review"

[4] A. A. Puntambekar "Data Structures Using C" First Edition 2009 pg no. 2-8 on "Selection and Analysis of Best Algorithm"

[5] Margaret H. Dunham, Sridhar "Data Mining: Introductory and Advanced Topics" First Edition - 2007 Vol. 1 Content - 6.3.2 Sampling technique in Apriori Algorithm