SYNOPSIS REPORT

on

RECOMMENDATION SYSTEM USING APRIORI ALGORITHM

Submitted by

Anshika Sharma (Enroll. No. R610217003)

Divyanshu Singh (Enroll No. R610217007)

Kaustavdeep Goswami (Enroll. No. R610217010)

Under the guidance of

Dr. Anurag Jain
Assistant Professor, Virtualization

SCHOOL OF COMPUTER SCIENCE
UNIVERSITY OF PETROLEUM & ENERGY STUDIES
Bidholi Campus, Energy Acres, Dehradun – 248007.

August - 2019

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

School of Computer Science

University of Petroleum & Energy Studies, Dehradun

Project Proposal Approval Form (2019-20)

Minor I

PROJECT TITLE: RECOMMENDATION SYSTEM USING APRIORI ALGORITHM

ABSTRACT

The world now stands on data. From morning till night each and every of our work is related to data and the data is stored in the data warehouse. Unfortunately, majority of our data remains unused. The trend has now changed. Today, most of the data is analyzed and processed to gain certain knowledge to profit the company.

Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue.

Data mining programs analyze relationships and patterns in data based on user's request. Data mining can be a cause for concern when a company uses only selected information, which is not representative of the overall sample group, to prove a certain hypothesis.

To sum up, data mining is the eye of our future, without it the world is blind.

Keywords: Data Mining, Apriori algorithm, Market-basket analysis, Association rule,

INTRODUCTION

Data mining is also known as data discovery and knowledge discovery.

The first step in data mining is gathering relevant data critical for business. Company data is either transactional, non-operational or metadata. Transactional data deals with day-to-day operations like sales, inventory and cost etc. Non-operational data is normally forecast, while metadata is concerned with logical database design. Patterns and relationships among data elements render relevant information, which may increase organizational revenue. Organizations with a strong consumer focus deal with data mining techniques providing clear pictures of products sold, price, competition and customer demographics.

For instance, the retail giant Wal-Mart transmits all its relevant information to a data warehouse with terabytes of data. This data can easily be accessed by suppliers enabling them to identify customer buying patterns. They can generate patterns on shopping habits, most shopped days, most sought for products and other data utilizing data mining techniques.

The second step in data mining is selecting a suitable algorithm - a mechanism producing a data mining model. The general working of the algorithm involves identifying trends in a set of data and using the output for parameter definition. The most popular algorithms used for data mining are classification algorithms and regression algorithms, which are used to identify relationships among data elements. Major database vendors like Oracle and SQL incorporate data mining algorithms, such as clustering and regression tress, to meet the demand for data mining.

Fig 1 : The Data Mining Cycle

PROBLEM STATEMENT

Once it is known that customers who buy one product are likely to buy another, it is possible for the company to market the products together, or to make the purchasers of one product the target prospects for another. If customers who purchase tomatoes are already likely to purchase onions, they'll be even more likely to if there happens to be an onion crate just beside the tomato aisle. By targeting customers, the effectiveness of marketing can be significantly increased – regardless of if the marketing takes the form of in-store displays, catalog layout design, or direct offers to customers. This is the purpose of market basket analysis – to improve the effectiveness of marketing and sales tactics using customer data already available to the company.

LITERATURE REVIEW

The efficiency of mining association rules is an important field of Knowledge Discovery in Databases. The Apriori algorithm is a classical algorithm in mining association rules. This paper presents an improved Apriori algorithm to increase the efficiency of generating association rules. This algorithm adopts a new method to reduce the redundant generation of sub-itemsets during pruning the candidate itemsets, which can form directly the set of frequent itemsets and eliminate candidates having a subset that is not frequent in the meantime. This algorithm can raise the probability of obtaining information in scanning database and reduce the potential scale of item sets. [1]

Market Basket Analysis(MBA) also known as association rule learning or affinity analysis, is a data mining technique that can be used in various fields, such as marketing, bioinformatics, education field, nuclear science etc. The main aim of MBA in marketing is to provide the information to the retailer to understand the purchase behavior of the buyer, which can help the retailer in correct decision making. There are various algorithms are available for performing MBA. The existing algorithms work on static data and they do not capture changes in data with time. But proposed algorithm not only mine static data but also provides a new way to take into account changes happening in data. [2]

Association Rules is one of the data mining techniques which is used for identifying the relation between one item to another. Creating the rule to generate the new knowledge is a must to determine the frequency of the appearance of the data on the item set so that it is easier to recognize the value of

the percentage from each of the datum by using certain algorithms, for example apriori. This research discussed the comparison between market basket analysis by using apriori algorithm and market basket analysis without using algorithm in creating rule to generate the new knowledge. [3]

Market-Basket Analysis is a process to analyze the habits of buyers to find the relationship between different items in their market basket. The discovery of these relationships can help the merchant to develop a sales strategy by considering the items frequently purchased together by customers. [4]

In this research, the data mining with market basket analysis method is implemented, where it can analyze the buying habit of the customers. Searching for frequent itemsets performed by Apriori algorithm to get the items that often appear in the database and the pair of items in one transaction. Pair of items that exceed the minimum support will be included into the frequent itemsets are selected. Frequent itemsets that exceed the minimum support will generate association rules after decoding. One frequent itemsets can generate association rules and find the confidence, which is uses a hybrid-dimension association rules. [5]

OBJECTIVES

Objective: To create Recommendation system based on Data Mining

- To generate patterns and relationships among data elements, render relevant information, which may increase organizational revenue.
- To analyze and draw conclusions about trends in consumer's marketing behavior.

METHODOLOGY

Phase 1

Requirement analysis

- Study concepts of Data Mining.
- Study of Association rule mining and Apriori Algorithm.

Phase 2

Designing and development

Figuring out a mathematically optimized algorithm which will be used to develop Recommendation system.

Designing and development is further divided into various phases.

This phase starts with the input from the Requirement and Analysis phase which will clarify the problem definition which will lead this project to the model development phase where a model will be created and further lead to the designing of the algorithm.

After completion of the designing of the algorithm phase, the focus will shift on the analysis of the algorithms and their implementation in this project.

Phase 3

Coding

On receiving system design documents, the work is divided into modules/unit and distributed among the team members and actual coding is started. Since in this phase the code is produced so it is the main focus of the developers. This is going to be the longest phase in this project.

The implementation of this project starts in terms of writing program in the suitable programming language and developing error free executable program efficiently, this phase primarily focuses on coding.

Implementation

- Creating a File(.txt) for data analysis.
- Implementation of Apriori algorithm
- Generating patterns and sets with Confidence value.
- Applying market-basket analysis using association rule.
- Analyzing and drawing conclusions of recommendation system with the required sets and patterns.

Phase 4

Testing

- Manual Testing

Phase 5

Integrating and implementing all the above phases.

SYSTEM REQUIREMENTS

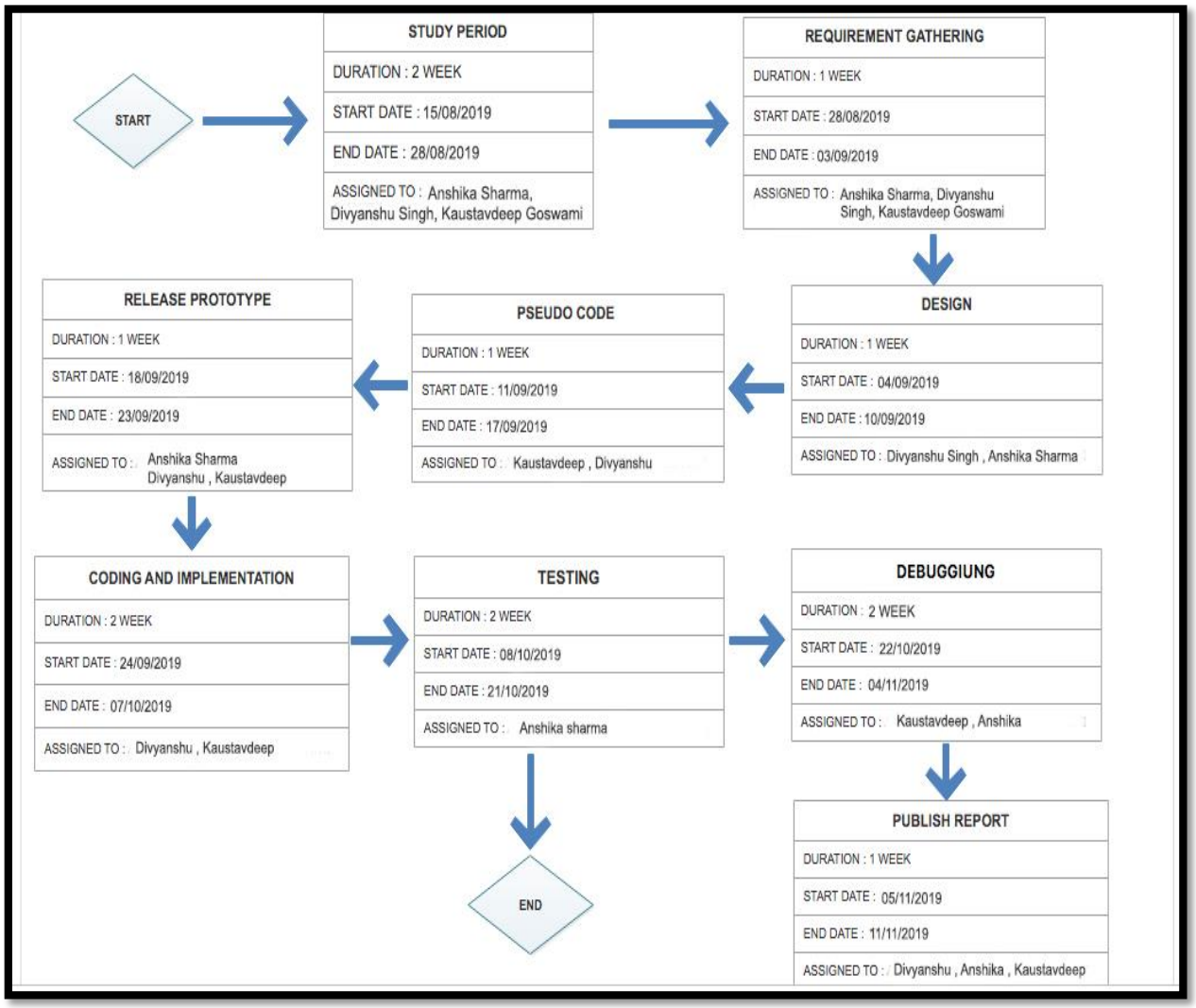
Software Requirements

Name of Component	Specification
Operating System	Windows 98 and above, Linux
Compiler	GCC

Hardware Requirement

Name of Component	Specification
Processor	Processor with speed of 500MHz
RAM	128 MB
Hard Disk	512 GB

SCHEDULE


START
STUDY PERIOD
DURATION : 2 WEEK
START DATE : 15/08/2019
END DATE : 28/08/2019
ASSIGNED TO : Anshika Sharma, Divyanshu Singh, Kaustavdeep Goswami
REQUIREMENT GATHERING
DURATION : 1 WEEK
START DATE : 28/08/2019
END DATE : 03/09/2019
ASSIGNED TO : Anshika Sharma, Divyanshu Singh, Kaustavdeep Goswami
DESIGN
DURATION : 1 WEEK
START DATE : 04/09/2019
END DATE : 10/09/2019
ASSIGNED TO : Divyanshu Singh , Anshika Sharma
PSEUDO CODE
DURATION : 1 WEEK
START DATE : 11/09/2019
END DATE : 17/09/2019
ASSIGNED TO : Kaustavdeep , Divyanshu
RELEASE PROTOTYPE
DURATION : 1 WEEK
START DATE : 18/09/2019
END DATE : 23/09/2019
ASSIGNED TO : Anshika Sharma Divyanshu , Kaustavdeep
CODING AND IMPLEMENTATION
DURATION : 2 WEEK
START DATE : 24/09/2019
END DATE : 07/10/2019
ASSIGNED TO : Divyanshu , Kaustavdeep
TESTING
DURATION : 2 WEEK
START DATE : 08/10/2019
END DATE : 21/10/2019
ASSIGNED TO : Anshika sharma
DEBUGGIUNG
DURATION : 2 WEEK
START DATE : 22/10/2019
END DATE : 04/11/2019
ASSIGNED TO : Kaustavdeep , Anshika
PUBLISH REPORT
DURATION : 1 WEEK
START DATE : 05/11/2019
END DATE : 11/11/2019
ASSIGNED TO : Divyanshu , Anshika , Kaustavdeep
END

REFERENCES

[1] Sheng Chai ; Jia Yang ; Yang Cheng, The Research of Improved Apriori Algorithm for Mining Association Rules, 2007 International Conference on Service Systems and Service Management Year: 2007 | Conference Paper | Publisher: IEEE , pp. 345-353.

[2] A.A. Raorane, R.V. Kulkarni, B.D. JitkarAssociation Rule – Extracting Knowledge Using Market Basket Analysis, Research Journal of Recent Sciences, 1 (2) (2012), pp. 19-27

[3] Patcharin Ponyiam, Somjit Arch-int, "Customer Behavior Analysis Using Data Mining Techniques", Application for Technology of Information and Communication (iSemantic) 2018 International Seminar on, pp. 549-554, 2018.

[4] Wan Faezah Abbas, Nor Diana Ahmad, Nurlina Binti Zaini, "Discovering Purchasing Pattern of Sport Items Using Market Basket Analysis", Advanced Computer Science Applications and Technologies (ACSAT) 2013 International Conference on, pp. 120-125, 2013.

[5]Md. Mahamud Hasan, Sadia Zaman Mishu, "An Adaptive Method for Mining Frequent Itemsets Based on Apriori And FP Growth Algorithm", Computer Communication Chemical Material and Electronic Engineering (IC4ME2) 2018 International Conference on, pp. 1-4, 2018.

[6] V. Saurkar Anand, V. Bhujade, P. Bhagat, A. KhapardeA Review Paper on various Data Mining Techniques
International Journal of Advanced Research in Computer Science and Software Engineering, 4 (4) (2014), pp. 98-101

Synopsis Draft verified by

Project Guide
(Name & Sign)

HOD
(Dept. of Systemics)