## ASSOCIATION RULE MINING

## B.Tech. Final Year Project Report

## BY

**PRATYAY ROY**



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## NEOTIA INSTITUTE OF TECHNOLOGY, MANAGEMENT & SCIENCE

## JHINGA, AMIRA, D.H.ROAD, 24-PARAGANAS (S) - 743368, WB (INDIA)

## November, 2016

## ASSOCIATION RULE MINING

## A Major Project Report

### Submitted in partial fulfillment of the requirements for the award of the degree

*Of*

#### Bachelor of Technology

*In*

##### **COMPUTER SCIENCE AND ENGINEERING**

## BY

|  |  |  |
| --- | --- | --- |
| **PRATYAY ROY** | **14400113021** | **131440110021** |

**Under the Guidance of**

Prof. Subrata Datta

## imgpsh_fullsize

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## NEOTIA INSTITUTE OF TECHNOLOGY, MANAGEMENT & SCIENCE

## JHINGA, AMIRA, D.H.ROAD, 24-PARAGANAS (S) - 743368, WB (INDIA)

## November, 2016

# imgpsh_fullsize

# CERTIFICATE OF ORIGINALITY

I hereby certify that the work which is being presented in the B.Tech. Final Year Project Report entitled **“ASSOCIATION RULE MINING”,** in partial fulfilment of the requirements for the award of the **Bachelor of Technology in Computer Science & Engineering** and submitted to the Department of Computer Science & Engineering of Neotia Institute of Technology, Management & Science, West Bengal is an authentic record of my own work carried out from July, 2016 to November, 2016 under the supervision of **Prof. Subrata Datta**.

The matter presented in this thesis has not been submitted by me for the award of any other degree elsewhere.

|  |
| --- |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  *Signature of Candidate* |
| **Pratyay Roy** |
| **14400113021** |
| **131440110021** |



**CERTIFICATE OF RECOMMENDATION**

This is to certify that the Project entitled “ASSOCIATION RULE MINING” has been submitted by **MS.** **PRATYAY ROY** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Science & Engineering of Neotia Institute of Technology, Management & Science, Jhinga, WB during the academic year 2016-2017.

|  |
| --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| *Signature of Supervisor* |
| Prof. Subrata Datta |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Head of the Department***

Department of Computer Science & Engineering

Neotia Institute of Technology, Management & Science, Jhinga, West Bengal

Date:

Place:



**ACKNOWLEDGEMENT**

This Project has been carried out to meet the academic requirements of Maulana Abul Kalam Azad University of Technology, West Bengal. I would like to put on record, my appreciation and gratitude to all who have rendered their support and guidance. Without them, it would not have been possible for me to shape this project.

I have received immense guidance from my project supervisor Mr. Subrata Datta, Professor and Head of the Computer Science Engineering Department, NITMAS. I would therefore like to convey my sincere gratitude to him.

**Pratyay Roy**

**INDEX**

**ABSTRACT**

The process of data mining produces various patterns from a given data source. The most recognized data mining tasks are the process of discovering frequent itemsets and frequent association rules. Numerous efficient algorithms have been proposed to do the above processes. Frequent pattern mining has been a focused topic in data mining research with a good number of references in literature and for that reason an important progress has been made. Association Rule mining (ARM) is one of the utmost current data mining techniques designed to group objects together from large databases aiming to extract the interesting correlation and relation among huge amount of data. In this article, we provide a brief review and analysis of the current status of frequent pattern mining and discuss a new objective measure to generate the rules.

**INTRODUCTION**

Data mining is a prominent tool for knowledge mining which includes several techniques: Association, Sequential Mining, Clustering and Deviation. It uses a combination of statistical analysis, machine learning and database management explore the data and to reveal the complex relationships that exists in an exhaustive manner. Additionally, Data Mining consists in the extraction of implicit knowledge (Previously unknown and potentially useful), hidden in large databases.

Data mining tasks can be classified into two categories: Descriptive mining and Predictive mining. Descriptive mining refers to the method in which the essential characteristics of the data in the database are described. Clustering, Association and Sequential mining are the main tasks involved in the descriptive mining techniques tasks. Predictive mining deduces patterns from the data in a similar manner as predictions. Predictive mining techniques include tasks like Classification, Regression and Deviation detection.

Mining Frequent Itemsets from transaction databases is a fundamental task for several forms of knowledge discovery such as association rules. An itemset is frequent if the subsets in a collection of sets of items occur frequently. Frequent itemsets is generally adopted to generate association rules. The objective of Frequent Item set Mining is the identification of items that co-occur above a user given value of frequency, in the transaction database. Association rule mining is one of the principal problems treated in KDD and can be defined as extracting the interesting correlation and relation among huge amount of transactions.

Formally, an association rule is an implication relation in the form X→Y between two disjunctive sets of items X and Y. A typical example of an association rule on "market basket data" is that "80% of customers purchasing bread also purchase butter". Each rule has two quality measurements, support and confidence. The rule X→Y has confidence c if c% of transactions in the set of transactions D that contains X also contains Y. The rule has a support S in the transaction. Efficient Analysis of Pattern and Association Rule Mining Approaches set D if S% of transactions in D contain XUY. The problem of mining association rules is to find all association rules that have a support and a confidence exceeding the user-specified threshold of minimum support (called MinSup) and threshold of minimum confidence (called MinConf ) respectively.