Using Enhanced AADMA Association Rule Mining Algorithm

On Automobile Dataset

A Dissertation Submitted

By

Gurpreet Batra

11311173

To

Department of Computer Science and Engineering

In partial fulfilment of the Requirement for the

Award of the Degree of

Master of Technology in Computer Science and Engineering

Under the guidance of

Ms. Alpana Vijay Rajoriya

(Assistant Professor, Lovely Professional University)

(MAY 2015)

PAC FORM

LOVELY PROFESSIONAL UNIVERSITY

Transforming Education, Transforming India

School of: Computer Science and Engineering

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the student	: Gurpreet Batra	Registration No	: 11311173
Batch	: 2013-2015	Roll No	: RK2306B68
Session	: 2014-2015	Parent Section	: K2306

Details of Supervisor:

Name	: Alpana Vijay Rajoriya	Designation	: Assistant Professor
UID	: 17447	Qualification	: M.Tech
		Research Exp.	: 1 year

Specialization Area: Database (pick from list of provided specialization areas by DAA)

Proposed Topics:-

1. An enhancement on Association Rule mining for a data set.
2. An analysis study in various Association rule mining techniques and their comparison.
3. An improvement on association rule data mining technique.

17447

Signature of supervisor

PAC Remarks: Topic 1 approved. Paper expected.

11057

11041 24/11/14

APPROVAL OF PAC CHAIRPERSON: Signature: Date:

*Supervision should finally encircle one topic out of three proposed topics and put up for an approval before Project Approval Committee (PAC).
*Original copy of this format after PAC approval will be retained by the student and must be attached in the Project/Dissertation final report.
*One copy to be submitted to supervisor.

ABSTRACT

One of the well-researched and most important techniques of mining data is Association Rule Mining. Association Rules as the name itself indicates includes finding correlations among sets of items in transaction database. As the enormous amounts of data is being collected and stored, more and more industries are taking a huge interest in mining such patterns from their databases. Discovering remarkable correlation relationships from huge amounts of transaction records in business helps in many decision making processes of business such as designing of the catalog, cross-marketing as well as the behaviour analysis of the customer shopping. Most famous algorithm of association rule mining is Apriori is used for knowledge discovery. Problems faced by Apriori algorithm includes the repetitive scanning of the database, generation of the large number of large candidate itemsets, ad adoption of only support. In our research work, we have developed a new algorithm named AADMA which is based on Apriori and it has removed many of the shortcomings of Apriori. The proposed work is based on automobiles study. The study shows that our developed AADMA provides the best effective result than the existing Apriori.

CERTIFICATE

This is to certify that **Gurpreet Batra** has completed M.Tech dissertation titled **Using Enhanced AADMA Association Rule Mining Algorithm On Automobile Dataset** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the dissertation proposal has ever been submitted for any other degree or diploma.

The dissertation is fit for the submission and the partial fulfillment of the conditions for the award of M.Tech Computer Science & Engg.

Date:

Signature of Advisor

Name: Alpana Vijay Rajoriya

UID :17447

ACKNOWLEDGEMENT

I would like to express my deepest gratitude towards my mentor **Ms. Alpana Vijay Rajoriya** for providing excellent guidance, advice, encouragement, supervision and inspiration throughout the development of this dissertation study. I would like to thank to the **Project Approval Committee members** for their valuable comments and discussions. I would also like to thank to **Lovely Professional University** for the support on academic studies and letting me involve in this study.

And last but not least, I find no words to acknowledge the moral support rendered by my parents and friends. All this has become reality because of their blessings and above all by the grace of **GOD.**

GURPREET BATRA

(11311173)

DECLARATION

I hereby declare that the dissertation proposal entitled **"Using Enhanced AADMA Association Rule Mining Algorithm On Automobile Dataset"** submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date: **Investigator**

Reg. No. 11311173

TABLE OF CONTENTS

SR. NO.	TOPIC	PAGE NO.
1	**INTRODUCTION**	**1**
	1.1 Database	1
	1.2 Data Mining	1
	1.3 Data Mining Techniques	3
	1.4 Types of Data	3
	1.5 Association Rule Mining	4
	1.5.1 Properties of Association Rule Mining	6
	1.6 Apriori Algorithm	6
	1.6.1 Flowchart of Apriori Algorithm	7
2	**REVIEW OF LITERATURE**	**11**
3	**PRESENT WORK**	**16**
	3.1 Problem Formulation	16
	3.2 Scope of the Study	17
	3.3 Objectives	17
	3.4 Research Methodology	18
	3.4.1 Working of Proposed Technique	18
	3.4.2 Flowchart of Proposed Technique	19
	3.4.3 Algorithm of Proposed Technique	20
4	**RESULTS AND DISCUSSIONS**	**21**
	4.1 Tools Used	21
	4.2 Results and Discussions	23
5	**CONCLUSION AND FUTURE SCOPE**	**45**
	References	46
	Appendix	48

LIST OF TABLES

TABLE NO.	TOPIC	PAGE NO.
1	Transactional Data	9
2	Minimum Support used by Apriori and AADMA	39
3	Minimum Confidence used by Apriori and AADMA	40
4	Reduction in number of cycles by AADMA as compared to Apriori	41
5	Reduction in set of Large Itemsets by AADMA as compared to Apriori	42
6	Best Rules generated by Apriori and AADMA	43

LIST OF FIGURES

FIGURE NO.	TOPIC	PAGE NO.
1	Data Mining, a phase in knowledge discovery	2
2	Market Basket Analysis	5
3	Frequent Itemset generation	8
4	Strong Association Rule generation	9
5	Generating C1 and L1 itemsets	10
6	Generating C2 and L2 itemsets	10
7	Candidate itemsets and frequent itemsets generation	10
8	Flowchart depicting working of AADMA	19
9	Weka Interface	21
10	Weka Explorer	22
11	NetBeans Interface	23
12	Automobile Dataset	24
13	Automobile Dataset opened in Weka Explorer	25
14	Visualization of Origin attribute of Automobile Dataset	25
15	Visualization of Man.Trans.Avail attribute of Automobile Dataset	26
16	Visualization of Drive Train attribute of Automobile Dataset	27
17	Visualization of Airbags attribute of Automobile Dataset	27
18	Visualization of Type attribute of Automobile Dataset	28
19	Visualization of Model attribute of Automobile Dataset	29
20	Visualization of Country attribute of Automobile Dataset	29
21	Visualization of Name attribute of Automobile Dataset	30
22	Visualization of Payment Type attribute of Automobile Dataset	31
23	Visualization of Manufacture attribute of Automobile Dataset	31
24	Selection of Apriori Algorithm	32
25	Run information of Apriori	33

FIGURE NO.	TOPIC	PAGE NO.
26	Results of Apriori on Automobile dataset	34
27	Rules generated by Apriori on Automobile dataset	35
28	Log information of Apriori on Automobile dataset	35
29	Selection of AADMA	36
30	Run information of our AADMA	37
31	Results of AADMA on Automobile dataset	37
32	Rules generated by AADMA on Automobile dataset	38
33	Log information of AADMA on Automobile dataset	39
34	Minimum Support used by Apriori and AADMA	40
35	Minimum Confidence used by Apriori and AADMA	41
36	Reduction in number of cycle performed by AADMA as compared to Apriori	42
37	Reduction in set of Large Itemsets by AADMA as compared to Apriori	43
38	Best Rules generated by Apriori and AADMA	44

CHAPTER 1
INTRODUCTION

1.1 Database

In the operation as well as management of an organization data is a very precious resource. So organizing data in a meaningful way is of utmost important. Data that is not organised has no meaning at all. Increased amount of data in various areas like banks, companies, universities, healthcare sector, etc has led to storing data in an organised way so that we can use it again and again. Data management is a discipline and the primary focus of it is to organise the data. To manage data efficiently typically computer database is used. The size of database can range from few megabytes to several terabytes. Data present in a database is related and is shared between several users as well as applications. Operations that are commonly performed on databases includes insertion operation, deletion operation, updation operation and selection operation. Databases needed to be managed and for managing them a software package known as Database Management System (DBMS) is used. All of the operations on databases are performed by it. It basically does the job of an intermediator between user and database. Database technology after the development of database systems moved further and led to the development of next technologies. These includes advanced database systems, data warehousing, and data mining for doing advanced data analysis. [6]

1.2 Data Mining

Process of finding advantageous patterns from gigantic data is known by the name of data mining. As the information technology has expanded and developed, an extremely large quantity of data has been generated in many different areas. With the increase in the research in the field of the databases and in the field of information technology has enabled us to store this data as well as to manipulate this precious data for making of further decisions. Fundamentally, mining of data is a process which extracts the precious information as well as potential patterns among gigantic amount of data. It is also known as the process of discovering knowledge from data. Basically, in this process of mining data, analyses of information from different perspectives and summarizing it into helpful information for fund raising and trimming costs. Data mining software has been used as an analytical tool for analysing of data and providing it to users for examining it from different angles, providing facility of constructing categories and concluding relationships. The main aim in process of data mining is mining information from a dataset and transforming it into structure that are easily understandable. Converting data into knowledge helps in making quality decisions and process is known as knowledge discovery. [5]

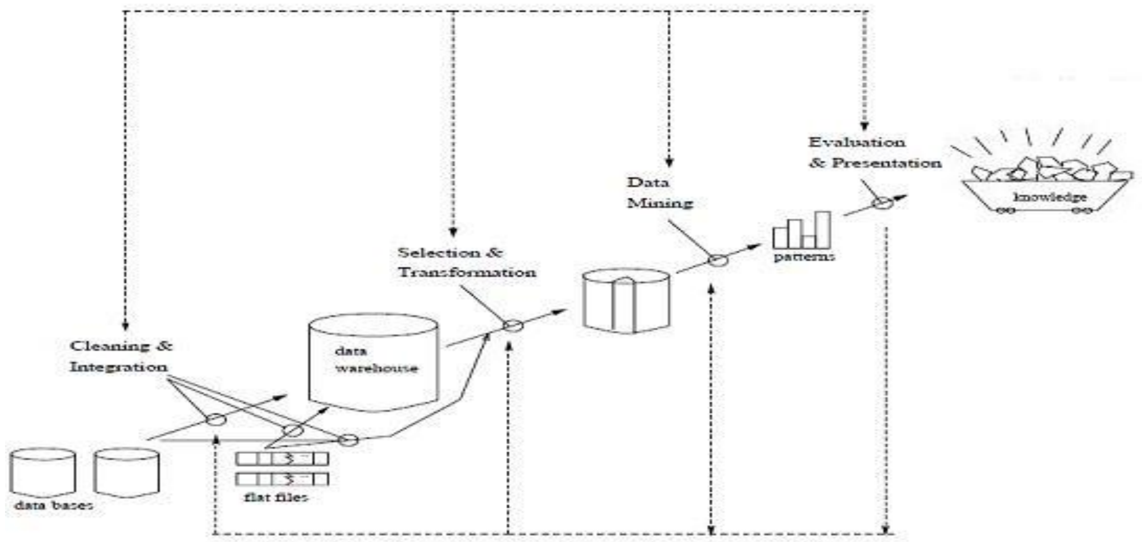

Figure 1. Data Mining, a phase in knowledge discovery

Steps present in knowledge discovery are:

1. Data cleaning that includes removing data that is not consistent and noisy data.

2. Data integration which involves integrating data from multiple data sources.

3. Data selection in which only task relevant data is extracted from the datawarehouse.

4. Data transformation that involves transforming the data into relevant forms by performing aggregation and summarization operations.

5. Data mining step involves application of intelligent methods for extracting data patterns.

6. Pattern evaluation involves identifying the truly interesting patterns that represent knowledge.

7. Knowledge presentation step involves the usage of techniques of knowledge representation as well as of visualization for presenting the knowledge so mined to users.

There are numerous application of data mining —it has been utilized seriously and also widely by marketers, for direct marketing, cross-selling or up-selling; by financial institutions, for credit scoring and fraud detection; by manufacturers in quality control, maintenance scheduling; and by retailers, for market segmentation and store layout.

1.3 Data Mining Techniques

Discovering knowledge from databases includes usage of different techniques and algorithms such as Association rules, Classification, Predictions, Decision trees, Artificial intelligence. These techniques also includes neural networks, clustering and neural networks. Advanced data mining techniques are used to discover relationships and other hidden patterns. Association rule mining is among the important data mining applications. Our research is focused on Association rule mining.

1.4 Types of Data

Techniques of data mining can be applied to following different types of data:

1. Relational database:

It contains form of relations known as tables that contains tupples(rows) and attributes(columns) that contains data.

2. Transactional database:

It has data in transaction form. Each transaction is basically a record which has its own unique transaction id. Each transaction has a transaction id and list of items that are purchased in that transaction.

3. Spatial database:

This includes data of maps. It stores data representing objects in geometric space.

4. Temporal and time-series database:

Temporal database includes data such as stock exchange data that changes rapidly and time-series database includes biological sequence data, heartbeat of patients, etc.

5. World-Wide Web:

It includes hypertext, audio, video, text data. It is a widely distributed repository of information that is made available by internet.

1.5 Association Rule Mining

One of the glaring, well explored technique in mining data is named as Association Rule Mining. As the name itself explains, it includes investigating associations in the sets of items in database of transactions. In relational database, dominant association rules can be pulled out from the data, otherwise unrelated. These rules have the format of IF- THEN form. The main idea behind the process of ARM is to investigate dataset for frequent patterns. Frequently occurring patterns are named as Frequent Patterns. Instance of it would be like "If someone buys a bottle a printer then he is 83% likely to buy a pack of papers".

An antecedent and consequent, the two components of association rules. Antecedent represents the IF part whereas THEN represented by Consequent. Antecedent is the item which is found in the data, whereas the item which is found in combination with the antecedent is consequent.

$$\text{Antecedent} \longrightarrow \text{Consequent}$$

$$\text{Support } (A \longrightarrow B) = P\ (A \text{ and } B)$$

$$\text{Confidence } (A \longrightarrow B) = P\ (A \mid B)$$

$$= \frac{\text{Support } (A \longrightarrow B)}{\text{Support } (A)}$$

The rules for association are made by identifying the repeated if/then patterns. Parameters named Support, Confidence are used to detect the most vital relationships. Support is used to indicate the most frequently appearing items. Confidence represents the figure of if/then statements that have been declared true.

Association Rule Mining is normally seen as a two-step process:

- First step is to find all of the frequent itemsets. Frequent itemsets includes only those itemset that has satisfied minimum support threshold.
- Second step is the generation of strong association rules from frequent itemsets. Only those rules are considered that satisfies the minimum support threshold as well as minimum confidence threshold.

Association rules are generated. Support, confidence is calculated. Threshold for minimum support as well as of minimum confidence is assumed. Only rules satisfying both the thresholds are considered to be true otherwise they are considered to be false. Accuracy is acknowledged as the confidence. And coverage is regarded as support. Accuracy implies the probability of precedent to be true in case the antecedent is true. Existence of high accuracy implies that this is the rule which is extremely dependable. The number of records on which the rule applies to is indicated by coverage. A rule that has high coverage is the rule that is used very regularly. Important role is played by these rules in areas like data analysis of shoping basket, store layout, product combinations and designing catalogue.

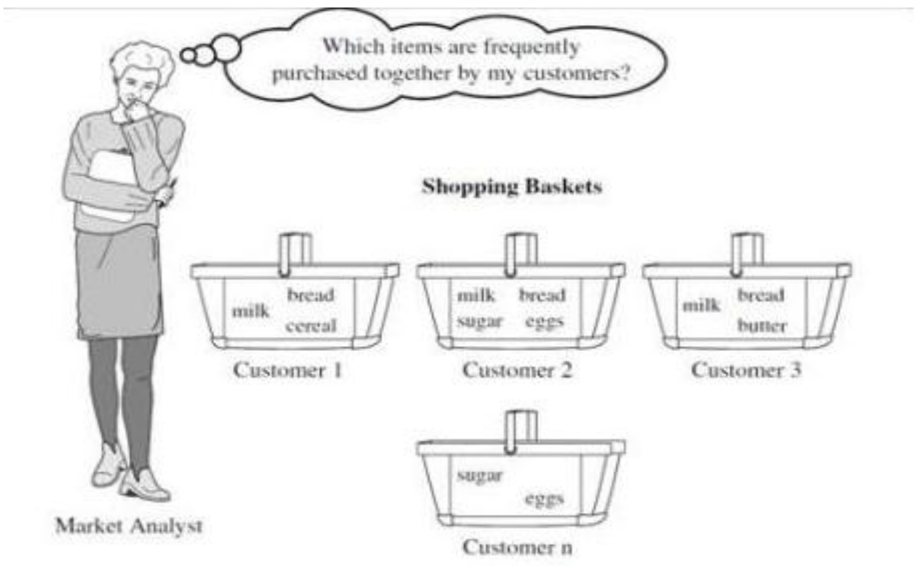

Figure 2. Market Basket Analysis

1.5.1 Properties of Association Rule Mining

1. Expresses the relationship between items or objects and how they tend to group together.

2. Simple to understand.

3. Provides useful information.

4. Data is considered transactional or relational.

Example: In an online mobile store they always give some suggestions after you buy a mobile, for instance, once you buy a mobile, a list of related accessories such as: Flip covers 40%, Smart watch 25%, will be presented to you as further purchasing recommendation.

In above example, association rules are: when the mobile is bought, 40% of the time the Flip cover is bought together, and 25% of the time the Smart watch is bought together. The rules discovered from the transaction database of the mobile store can be very helpful to rearranging the way of how to place those related accessories. These rules helps the store manager to create his market strategies such as: by promotion of the mobile, it can increase up the sales of the other two accessories mentioned in the example. Association rules are extensively employed in various areas like risk management, inventory control, market, etc.

1.6 Apriori Algorithm

Apriori, for Boolean association rules, is the conventional worldwide used algorithm for mining frequent patterns. Beforehand knowledge of frequently occurring itemset's properties is used by the algorithm therefore it is called as Apriori, made for executing transactional databases. It employs categorical attributes, implementing "bottom up" approach, extending one item at an instant of frequent subset. Termination of algorithm occurs on non-successful extensions. Chief motive of the algorithm is to find associations among different data sets. Transaction contains number of items which constitutes a set of data. Strong association rules outputted by algorithm, tells us the frequency of items in transactions. Algorithm hires Apriori property that states "Any subset of frequent itemsets must be frequent”. This helps in trimming search space. Iterative tactic is used by algorithm in which (k+1)-itemsets are explored from k- itemsets. First of all, commonly occurring itemsets i.e. 1-itemsets (C1) is generated by looking over the database for occurrence count of each item. Then, only items that satisfies the threshold of minimum support in C1 are taken for set L1. L1 is used for performing join operation in which L1 is joined with itself to produce candidate set of 2-itemsets i.e. C2. The itemsets that holds the support threshold are taken for L2. Generation of candidate(C) itemsets, L itemsets comes to an end when no more frequent k-itemsets are found.

Thus, it is broken in a two-step process which is employed to explore frequent itemsets: join and prune actions.

- **Join:**

Joining LK-1 with own, generates candidate k-itemset (CK).

- **Pruning:**

Ck, a subset of Lk, might contain members that may not be frequent. Although all frequent k- itemsets are its members. For trimming the Ck’s size Apriori property is employed. The Ck candidate’s are counted by a database scan which leds to the determination of Lk. Therefore, candidates that holds the support threshold are retained in Lk.

1.6.1 Flowchart of Apriori Algorithm

- **For Finding Frequent Itemsets:**

As shown in below flow chart, first of all dataset will be loaded and parameters support and confidence will be set. The decision maker will specify the minimum support and confidence threshold in advance. From the dataset, first of all the candidate set will be built and it will contain all the items. In this iteration the support count of all the items will also be calculated. Then from the candidate set, frequent itemset will be generated which contains all items having support greater than minimum support. From this frequent itemset next candidate itemset is generated by joining the frequent itemset with itself and from that candidate set the next level of frequent set is generated. This process will be continued until the frequent itemset generated turns out to be null. The last obtained itemset will be chosen for making association rules.

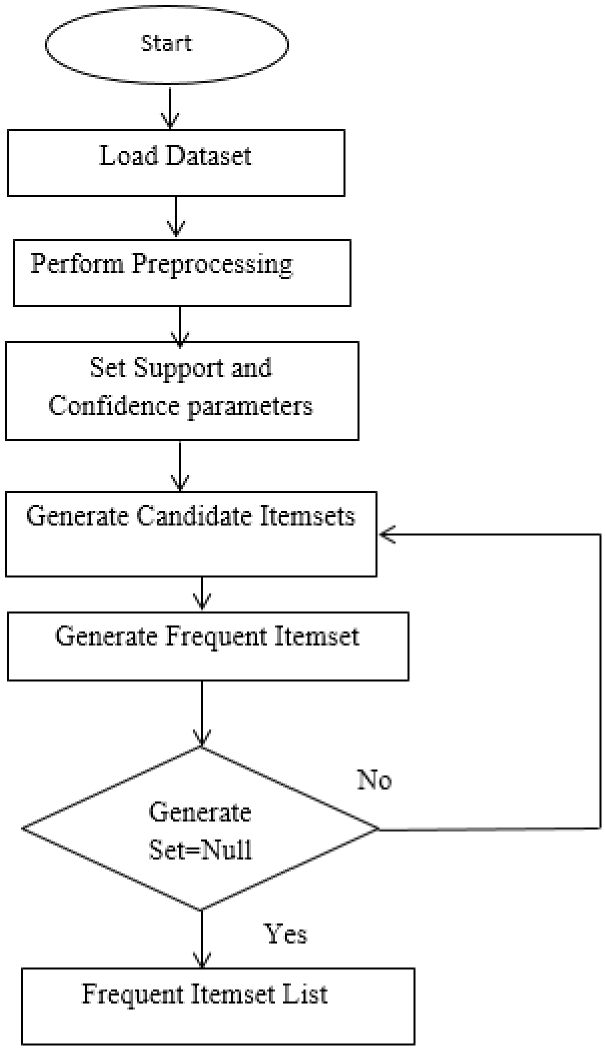

Figure 3. Frequent itemset generation

- **Generate Strong Rules**:

When association rules are generated first of all the confidence of the rules will be checked against the minimum confidence threshold. Then if it passes that then that rule will be outputted otherwise if the rule does not satisfy the minimum confidence threshold then it will be discarded.

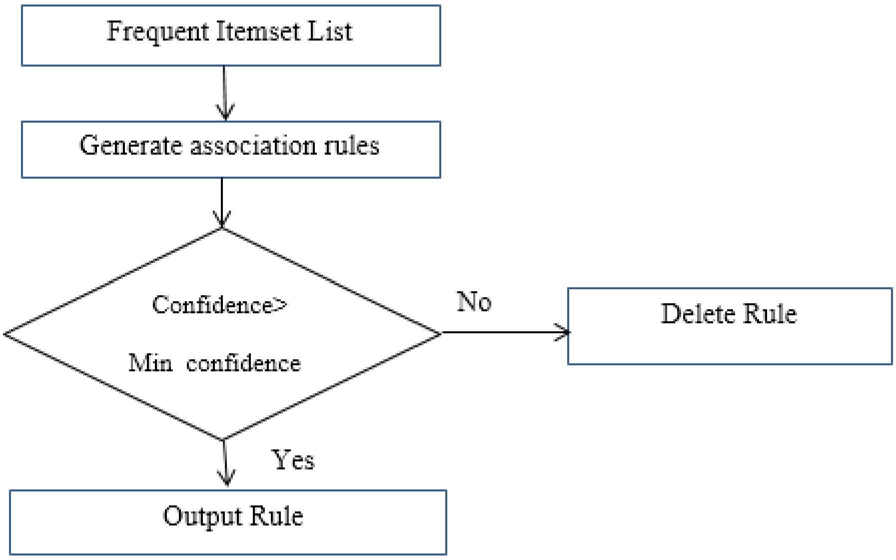

Figure 4. Strong Association Rule generation

Example: Let us consider a transaction database as shown in following table:

Min_Sup=2, Min_conf=70%

TID	ITEMS
T1	A,B,E
T2	B,D
T3	B,C
T4	A,B,D
T5	A,C
T6	B,C
T7	A,C
T8	A,B,C,E
T9	A,B,C

Table 1. Transactional Data

Apriori algorithm is employed as follows to find the frequent itemsets:

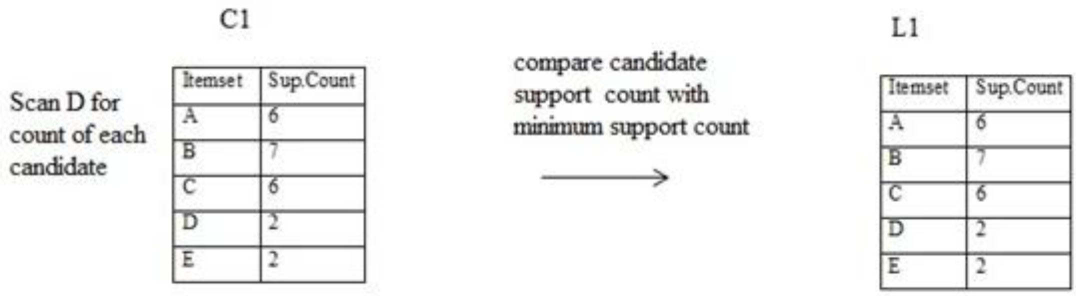

Figure 5. Generating C1 and L1 itemsets

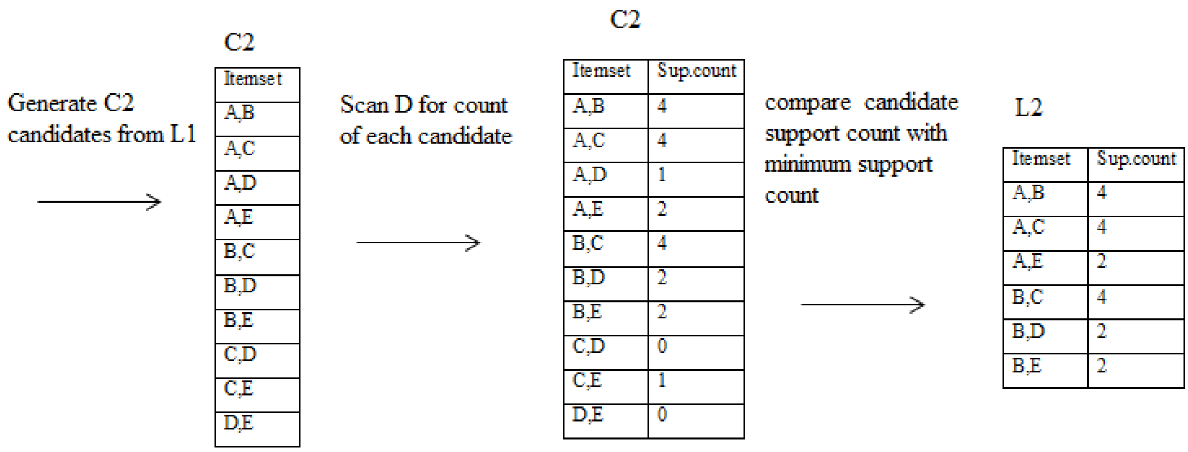

Figure 6. Generating C2 and L2 itemsets

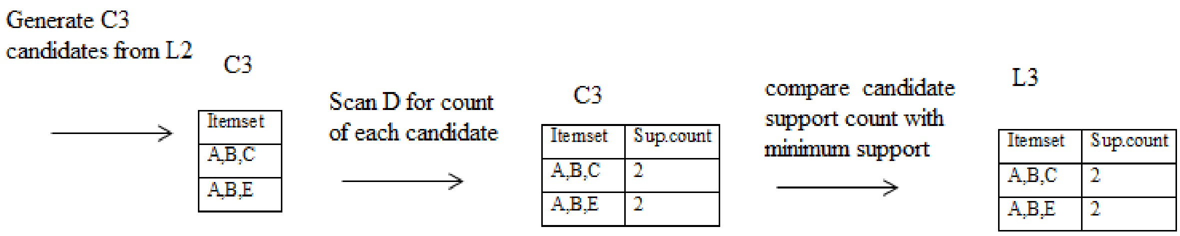

Figure 7. Candidate itemsets and frequent itemsets generation

After getting the frequent itemsets, the association rules are formed and those rules satisfying the minimum confidence threshold are known as strong association rules.

CHAPTER 2

REVIEW OF LITERATURE

Gitanjali J et al (2014) A large amount of data from various perspectives is analyzed in data mining which generates useful information. This information is very helpful in knowing history and predicting the future trends. A health care sector in present era generates a huge amount of data that includes information about diseases, electronic patient details, diagnosis methods, etc. This data is not mined so the important information remains hidden. So in this paper, association rule mining technique of data mining is used to mine important and significant knowledge that can be used by the healthcare administrators to improve their service quality. Here, apriori algorithm is applied to find frequency of diseases in a given period of time for a particular geographical location and it is found that patients are affected by 4 different diseases during a year at four geographical areas. [4]

Mohammed Al-Maolegi et al (2014) Knowledge can be discovered from association rules which are generated through Apriori Algorithm. However classical apriori algorithm involves scanning of whole database again and again. This leads to wasting of time. In this paper, an improvement on apriori is presented. In this, all of the database is not scanned repetitively. Instead, while generating the frequent 1- itemset from candidate set, only those items with their support count are recorded which satisfies the threshold of minimum support and also the transaction id of the frequent items is recorded. With this, the need for searching the items in whole database is eliminated. Thus, allowing for conducting their search only in the transactions whose ids are already stated. For producing the next itemset that is frequent 2- itemset, the join operation is performed on frequent 1- itemset by joining it with itself. The same process of recording of the transaction id as well as performing scan in between only those transactions is repeated for finding all of the k frequent itemsets. Thus, the paper provides a technique of reducing the operation of scan of large number of transactions in database by only performing the scan of the transactions specified in the transaction id. [9]

Rupinder kaur et al (2014) In this paper, association rule mining algorithms have been introduced. It is an important technique helping in finding out the relationships between data. Antecedent and consequent are two parts of these rules. Two basic measures are support, confidence. Different approaches that are used with apriori algorithm are discussed. It includes apriori improve which deals with the time and space concern of basic apriori using hash table and hash table and generating L2 directly from one scan. Another approach named Capriori used to set the item code to 1 if it is present in the transaction otherwise set to 0. Then counting support on this basis to generate the L1. Then L2 is produced by performing AND operation on L1. Thus all frequent sets are produced in this way and multiple scan need not to be performed. Another approach is Probability Apriori Multiple minimum support approach with uses multiple minimum support to discover frequent as well as rare itemsets. Another approach is weighted approach in which weights are also assigned to reflect importance of different items. Thus items having weighted support not satisfying the minimum weighted support were removed. [11]

U. C. Moharana et al (2014) provided an approach of using association rule mining in management of spare parts inventory. They have used the association rule mining to generate the associated spare parts which are used together in maintenance activities of the similar equipments. By using the association rule mining the spare parts dependencies are known and these helps in determination of the optimal spares kit based on the minimal cost. To check the correctness of the generated rules decision makers are used. Association rules so generated are only based on the binary information (0 or 1) of consumption records. These consumption records are taken from each maintenance activity. The advantage of genearating these rules includes inventory control, reducing storage space, buying all spares from single supplier at good discounts, and reduced labor cost.[12]

Chanchal Yadav et al (2013) proposes a approach to improve the traditional apriori algorithm. The new approach is based on the apriori but aims to reduce the drawback of it. The proposed algorithm consists of two phases. In the first phase, the cleaning algorithm is used to remove the noise or inconsistent or duplicate data. After data cleaning, preprocessing is done to remove the least associated item at the very early stage. In traditional apriori pruning is done but in the proposed approach pruning is not done because it uses preprocessing to do the similar task. The second phase includes the detection of frequent patterns by using improved association mining approach, only the filtered data will be used in this phase. Static groups are defined using the partition approach. Partitioning is done manually depending upon the appropriateness of the attributes. The proposed technique divided data into horizontal partitions .Support is computed in the same way as is done in the apriori algorithm. [1]

Dong Gyu Lee et al (2013) An association rule mining method is proposed that reduces the number of imprecise patterns produced by previous studies and provides interesting patterns. Interesting patterns discovered includes knowledge in Korean acute myocardial infarction registry. The proposed ARM method can discover target patterns associated with hypertension and diabetes from young AMI patients who are 45 years old or less. To reduce the imprecise associations between the antecedent and consequence of patterns, target patterns are defined and these target patterns are extracted in all patterns that are generated from frequent itemsets. Minimum support and confidence is specified in advance. As in association rule mining, target patterns are equal or greater than the minimum support and confidence thresholds. But only support and confidence cannot ensure interestingness of patterns. So combination of support and confidence with statistical measures such as lift, leverage and conviction to evaluate interestingness of target patterns proved sufficient. [2]

Jun Yang et al (2013) Equal importance was being given to all of the database transactions by traditional Apriori Algorithm. This results in the reduction of the importance as well as the reduction in the accuracy of the association rules so produced by it. In order to produce more sensible association rules level of importance given to different items should be different. In this paper, an improved algorithm named Feature Based Apriori is proposed which gets its base from the Apriori algorithm. Feautes are being used in improved algorithm. This means that every transaction is being provided with its own features that carries more information. In the traditional apriori all of the rules are mined. It includes no facility for looking for a particular one item. But, in this Feature Based Apriori, only transactions having the same features are scanned and computed for generating the association rules. Thus, it provides the association rules that are more sensible and reasonable. Improved algorithm is studied under the Book Recommendation System. Reader Type is the feature that is added to it. Thus, only the transactions having same type of reader are used to generate the strong association rules that are more sensible as well as reasonable. [7]

Marghny H et al (2013) In this paper, an approach is provided which is very helpful in solving the major problems of the ARM algorithms which are similar to Apriori. Some of the many problems of ARM algorithms includes the scanning the transaction database repeatedly, high memory dependency as well as a lot of computations involved in the generation of the candidate itemset. In this paper, solution to such problems is provided by using CountTableF1 and BinaryCountTableF1 algorithms. By using these, this method shows a huge and significant difference from the Apriori and the Apriori-like algorithms. The idea behind is to represent the transactions in the form of binary and decimal numbers. So it avoids the costly generation of candidate and test processing. Also the usage of subset and set identical properties becomes fast and simple. In this a new count table is constructed which is highly compact as well as considerably smaller when compared with the original database. In this the frequent itemsets are discovered using the Intersection. This operation is much faster when compared with the traditional item comparing method that is used in many algorithms that are Apriori-like. So the costly database scans are saved. The algorithm focuses on solving the candidate itemsets generation problem and support count problem so the need to generate the candidate itemset is not there. [8]

Kalyani M Raval (2012) In this paper, data mining techniques are introduced. Data mining is the process of mining advantageous and previously unknown information or patterns from huge databases. This information is then used in making very important and crucial decisions by decision makers. Various techniques of data mining that are discussed are association which includes the discovery of patterns based on relationship of item with other items. It is basically used to find the correlations among the items. Another technique of data mining is classification which is used to classify the items into different groups or classes which are predefined. Clustering is a technique of data mining which is used to group similar items such that items in one group are similar to each other but are dissimilar from items in another groups. The groups in clustering are not predefined. Prediction is another data mining technique which is used to predict the future based on the historical trends or patterns. Here, two types of variables are used: independent variables (attributes) which are already known and the response (dependent) variables which we want to predict. [10]

Farah Hanna AL-Zawaidah et al (2011) presents a association rule mining algorithm based on the conventional apriori algorithm. Conventional apriori when applied on large databases generates a huge number of association rules .This makes it very difficult for decision maker to decide which rules should be considered and which rules should be left. The main reason behind this was the iterative approach used by the apriori. Also there is a high as well as repeated disk access overhead which reduces the efficiency of the algorithm. In this paper, Feature Based ARM is proposed. Association rules are generated by using the features of the items and calculating the weight of candidate itemsets. The algorithm proposed encompasses the transformation of the database. It means it involves the reorganization of the transaction database into a matrix known as feature matrix. This greatly reduces the I/O accesses. It also fastens the process of mining data. Leverage measure is also used for reducing the candidate's itemsets. Thus, saving a memory to a great extent from storing the useless candidates. [3]

Yanxi Liu et al (2010) New computer technologies as well as database technologies are being developed at a rapid pace but these technologies are only for supporting the storage and quick retrieval of large scale database or data warehouses. For organizing this data and analyzing it data mining technologies are used. One the data mining technologies is Association rule mining that uses Apriori algorithm. The working of the Apriori algorithm is detailed and explained with the help of an example. Algorithm has used two sub-processes named Apriori-gen() which generates the candidates and applies apriori property to remove those subsets which are not frequent and, subset() which is used to do cumulative count for the generated candidates. The paper also highlights the shortcomings in the Apriori algorithm. [13]

CHAPTER 3

PRESENT WORK

3.1 Problem Formulation

Following are the problems faced by Apriori Algorithm:

1. Repetitive scanning of the whole database is required:

It is required to go over each transaction in the database for determining the support of candidate itemsets. This is required for determining whether the element is eligible to join the frequent itemset Lk. It is needed to scan database of transactions at least 10 times if 10 items are contained in the large frequent itemset. This causes an excessive input/output load.

2. Very Large candidate sets may be generated:

Huge number of candidate sets may still need to be generated. For instance, if frequent 1-itemsets are 10^4 then the 10^7 candidate 2-itemsets would be generated by the algorithm. Such large candidate sets are time and memory space challenges.

3. Only support is adopted:

Some affairs occurs very frequently while some affairs are very rare in real life. So the present algorithm has problem. Setting the minimum support threshold very high will led to less data coverage and significant rules may be left hidden. But on the other hand, setting minimum support threshold too low will generate a very huge number of rules and they may even include meaningless rules which would seriously reduce the rule efficiency as well as availability. So this will be misleading the decision making.

4. Narrow fitness landscape of the algorithm:

Only Boolean association rule mining is considered by algorithm. But, practical applications in present era may involve multi-dimensional as well as multi volume association rules. At this point, the present algorithm cannot be applicable generating the need of improving it.

3.2 Scope of the study

ARM is one of the data mining techniques that finds associations between antecedent and consequence. Representative algorithm that mining association rules is Apriori algorithm. Existing studies have been applied statistical methods and data mining approaches. Research study applied the existing apriori algorithm after enhancing it for finding important selling factors that affects the relevant sale of vehicles. Enhanced apriori algorithm provided more interesting rules than the rules that are provided by the existing algorithm. Research study shows that rules provided by the improved algorithm are better rules. Research work reduced computation time involved in the Apriori algorithm by improving the existing algorithm. Research study has improved the performance of existing algorithm by decreasing the computation time and the iterations required to generate rules.

3.3 Objectives

The problems taken for this research work is divided into some objectives which are as follows:

1. To reduce input output load.

2. To reduce Large candidate item set.

3. Traditional association rule mining approaches adopt an iterative technique to discover association rule, which requires very large calculations and a complicated transaction process. As, the existing mining algorithm cannot perform efficiently and the reasons being the high and repeated disk access overhead. Because of all these, proposed study will present improved apriori association rule mining approach that can efficiently discover the association rules in large databases.

4. To use data set related to multi-dimensional as well as multi-volume association rules.

5. To generate more meaningful and less number of rules.

3.4 Research Methodology

The approach developed is named as AADMA and it is based upon the existing Apriori algorithm of association rule mining. AADMA is an enhancement of the existing algorithm.

In this, a method has been developed which avoids the costly candidate generation as well as it also compresses the essential information about all the itemsets. This includes maximal length frequent itemsets, minimal length frequent itemsets. The method avoids expensive and repeated database scans. The proposed AADMA algorithm has major difference from Apriori and all of the other algorithms which are extended from Apriori. The main idea behind AADMA is in the way the transactions are represented. Here, transactions are represented in the values that are nominal and discrete. Thus, usage of subset and identical set properties have become very simple and at the same time very fast.

3.4.1 Working of Proposed Technique

Following are the phases of Proposed Technique:

- **Load Dataset**

At first, dataset is loaded. Dataset we have used is of automobiles data. Attributes of dataset includes Manufacturer name, Payment type, Car Name, Country, Model, Type, Airbags, Drive train, Manual Transition, Origin, Make.

- **Preprocessing**

The objective of this phase is to remove the noisy data. The data which can cause inconsistency or the missing values are removed from the dataset.

- **Applying AADMA**

After preprocessing phase, we apply the AADMA i.e. Advanced Association Data Mining Algorithm. Here, range for support is set as well as confidence parameter is set. With this, the number of scans are reduced. The candidate itemsets are generated in less number of iterations. The size of the large itemsets are also reduced. And, the rules so generated are also less in number.

3.4.2 Flowchart of Proposed Technique

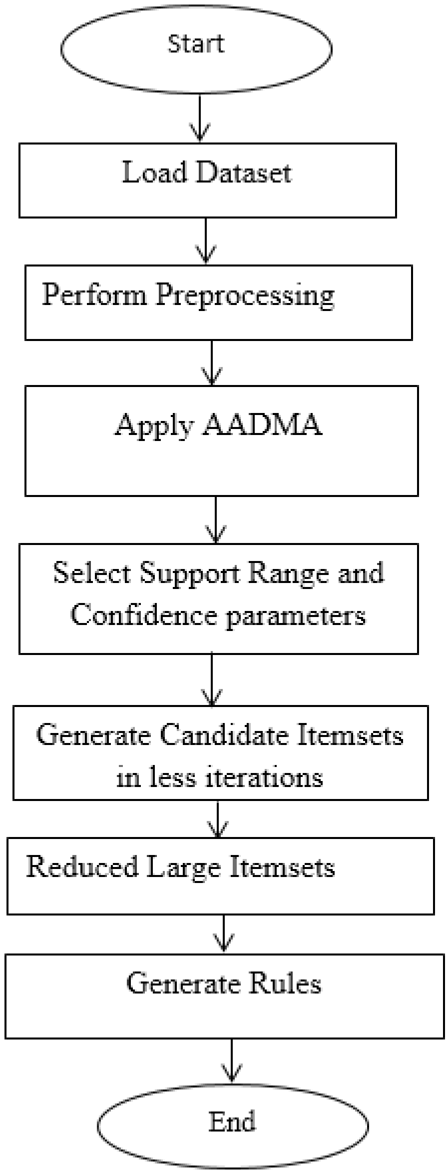

Figure 8. Flowchart depicting working of AADMA

As shown in the above flow chart, first of all our automobile dataset will be loaded. Then pre-processing of the dataset will be performed. This is performed with the objective of removing inconsistencies and to remove missing values. After this, our dataset is ready for the application of the AADMA algorithm. During this, the range for the support is provided and parameter minimum confidence is determined. By doing so, the dataset need not to be scanned again and again. The candidate itemsets that are produced comes out in very less number of iterations. As the number of iterations are reduced so the time taken is also reduced. The size of large itemsets is reduced. The rules that comes out are also less in number as compared to the classical algorithm.

3.4.3 Algorithm of Proposed Technique

AADMA, the improved Apriori algorithm is described in following steps:

Input: D_i, a database of transaction, Min_sup: the minimum support count threshold

1. In the very first iteration of the algorithm the C_i (set of candidate itemset 1) is generated. Each item is a member of C_i. All of the transactions are scanned to count the number of occurrences of each item.
2. L_i, the set of frequent item sets is then determined by comparing the count of the candidate with the minimum support count. L_i contains C_i itemsets satisfying minimum support.
3. To generate the L_{i+1}, the set of frequent 2-itemsets, the algorithm generates a candidate set of 2-itemsets C_{i+1} and then the transactions in D are scanned and the support count of each candidate item set in C_{i+1} is accumulated and then L_{i+1} is generated from C_{i+1} which satisfies the minimum support.
4. Then D2 is determined from L_{i+1}.
5. Generate C_{i+2} candidates from L_{i+1} and scan D2 for count of each candidate and then selecting those for L_{i+1} which satisfies the minimum support.
6. At the end of the pass, determine which of the candidate item sets are actually large, and those become the seed for the next pass.
7. This process continues until no new large item sets are found.
8. From these Large itemsets, rules are generated and those rules which satisfies the minimum confidence are selected.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Tools Used

- **Weka:**

The tool used by us is Weka Data Mining Software. It contains machine learning algorithms which are used for the tasks of data mining. These algorithms can be applied directly onto the datasets. Weka also provides the facility to call our own Java code. It contains tools which helps us in performing pre-processing of data. We can perform classification, regression, clustering of the data using Weka tool. It also provides us with tools for generating Association rules and Visualization of the datasets.

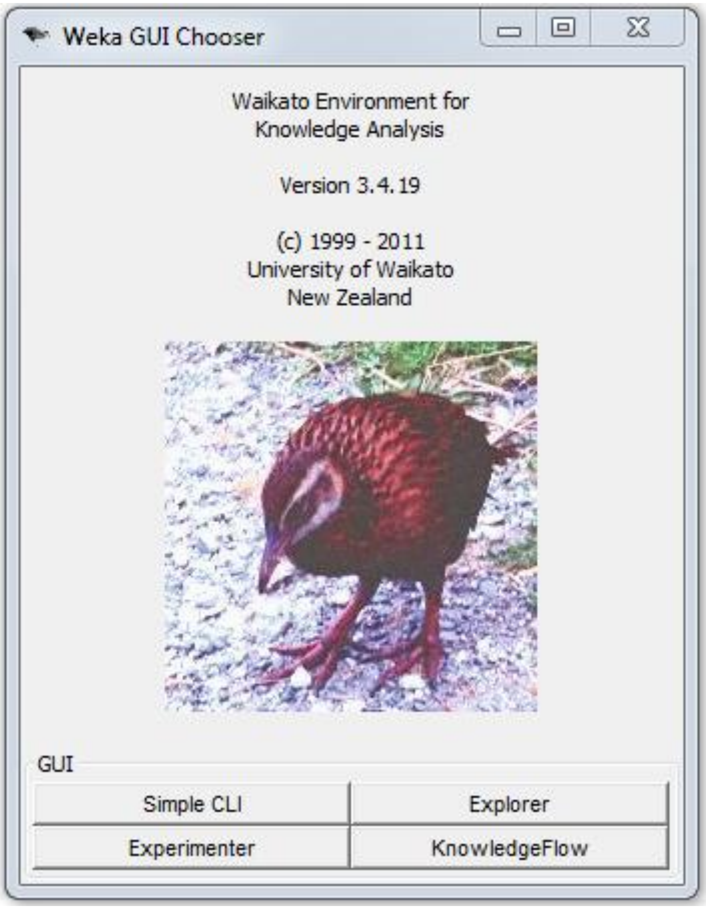

Figure 9. Weka Interface

When we run Weka, the above shown interface will appear. Here we select the Explorer.

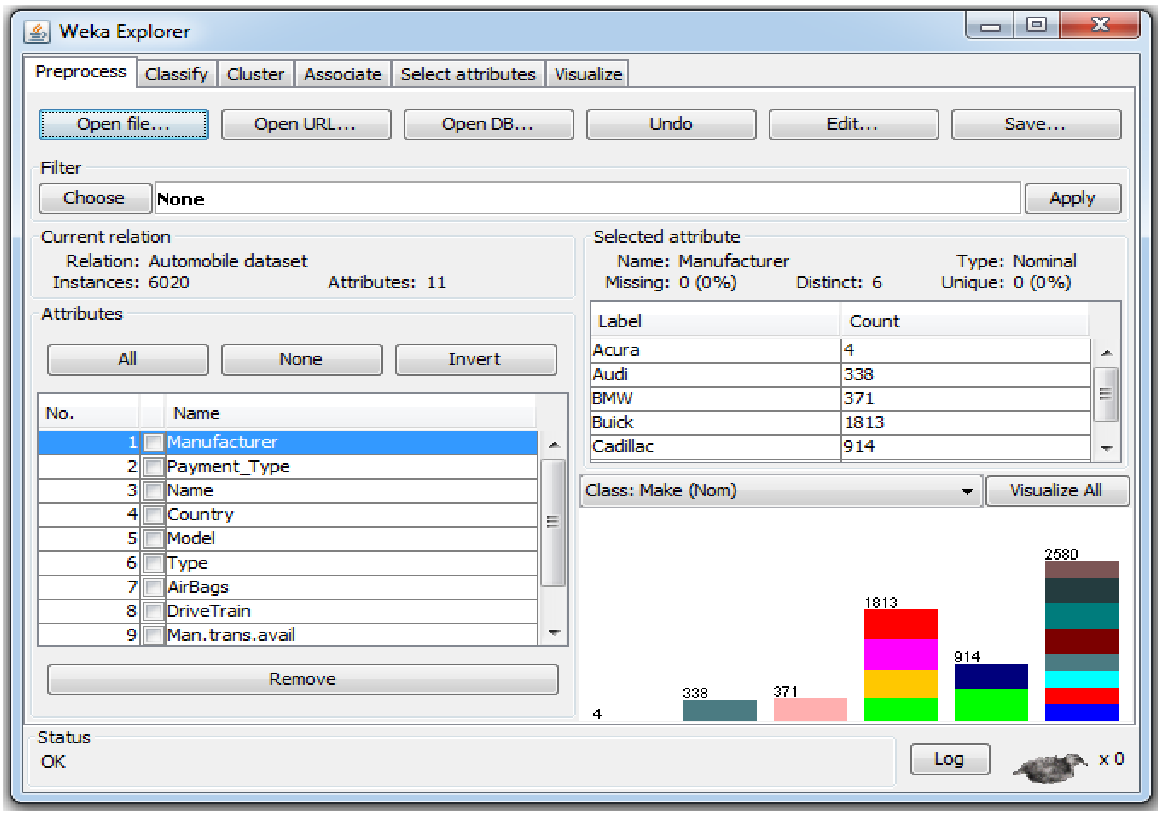

Figure 10. Weka Explorer

The explorer will open the above shown interface. Here, we can import our dataset which can be a excel file or ARFF format by clicking on 'Open File' tab. This will show our dataset in visual form as shown in the figure. We can select different attribute and can accordingly see their visualization. Different tabs present in this interface are Preprocess, Classify, Cluster, Associate, Select Attribute and Visualize. Preprocess contains filters which help in performing preprocessing of the dataset. Classify contains the algorithms for performing classification of dataset. Associate contains algorithms for the association rule mining tasks. We can embed our own developed java code in this also. In this research, we have used Weka 3.4 version.

- **NetBeans:**

NetBeans is platform which is written in java and provides us facility for the software development. This platform allows us to develop the application from modules. The applications that are based on the NetBeans Platform, including NetBeans IDE can also be extended by third party developers. It is a cross-platform. This means that it can run on Microsoft Windows as well as on Mac Os X. It can also run on Linux, Solaris and many other platforms that supports the compatible JVM. One of the main features of this platform are user interface management by providing menus, toolbars. For Java Programming Language NetBeans IDE has built in support. It is an open source which is freely available.

It allows us to do the coode editing which is fast and smart. NetBeans Editor indents the lines as well as matches brackets and words, highlights source code both semantically and syntactically. Project management is very easy and efficient in NetBeans. The version of NetBeans used in our research work is 8.0.2

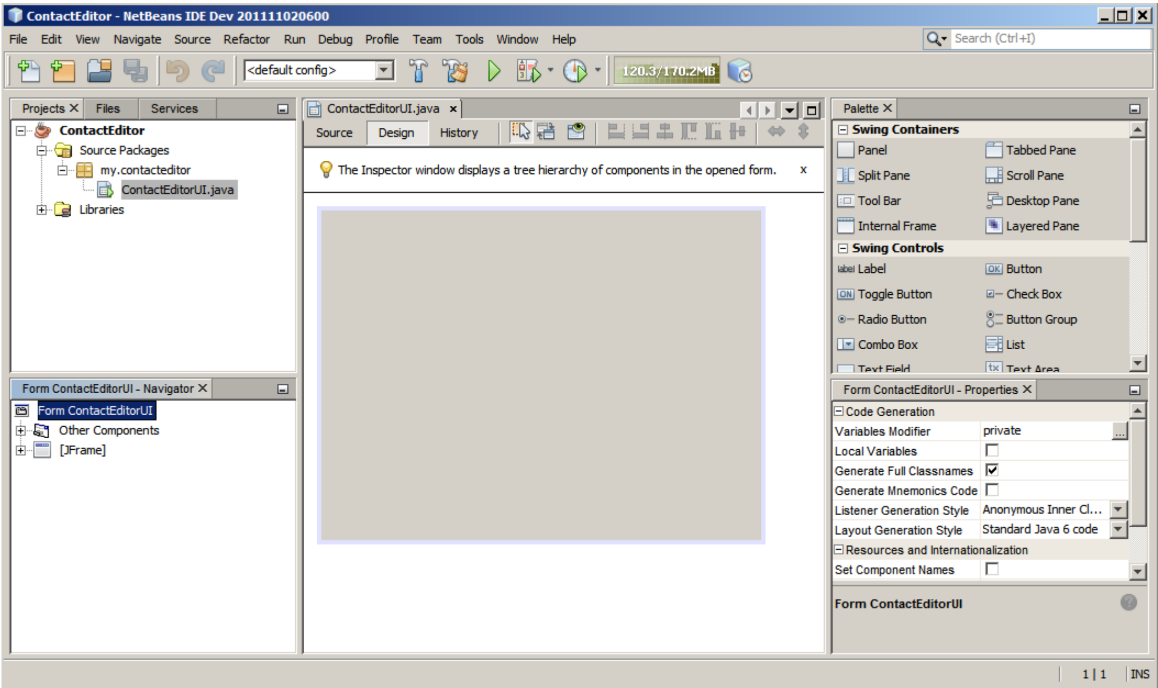

Figure 11. NetBeans Interface

4.2 Results and Discussions

In order to test our methodology, we applied our enhanced AADMA and Apriori on our Automobile dataset and found the following results:

1. Efficiency has been increased.

2. Traditional Apriori involves 14 iterations. The number of these iterations are reduced to 11 in AADMA.

3. AADMA has less computation time as compared to existing Apriori.

4. Size of large itemsets have been reduced to 3 using AADMA which earlier was 4 with existing Apriori..

5. Number of rules produced by AADMA are less as compared to existing Apriori. With existing Apriori 10 rules are produced but with AADMA 8 rules are produced.

Our aim was to optimize those phases of Apriori which consumes a lot of time and these includes the phase in which the small itemsets are counted that is the initial iterations. The main enhancements involves the use of an advanced method for storing candidate itemsets as well as for counting their support. It has used effective pruning methods which wisely reduce the dataset size as execution progresses. A comprehensive performance study using Automobile dataset has been done. Results shows that our technique is efficient as well as scalable when compared with other methods.

- **Dataset Used**

The dataset we have considered is of Automobiles Data. It contains the attributes named Manufacturer name, Payment type, Car Name, Country, Model, Type, Airbags, Drive train, Manual Transition, Origin, Make.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Manufactu	Payment_	Name	Country	Model	Type	AirBags	DriveTrair	Man.trans	Origin	Make	
2	Acura	Mastercar	carolina	United Kir	Integra	Small	None	Front	Yes	non-USA	Acura Integra	
3	Acura	Visa	Betina	United Sta	Legend	Midsize	Driver & P	Front	Yes	non-USA	Acura Legend	
4	Audi	Mastercar	Federica e	United Sta	Legend	Compact	Driver onl	Front	Yes	non-USA	Audi 90	
5	Audi	Visa	Gouya	Australia	Legend	Midsize	Driver & P	Front	Yes	non-USA	Audi 100	
6	BMW	Visa	Gerd W	United Sta	Century	Midsize	Driver onl	Rear	Yes	non-USA	BMW 535i	
7	Buick	Visa	LAURENCE	United Sta	Century	Midsize	Driver onl	Front	No	USA	Buick Century	
8	Buick	Mastercar	Fleur	United Sta	LeSabre	Large	Driver onl	Front	No	USA	Buick LeSabre	
9	Buick	Mastercar	adam	United Sta	Roadmast	Large	Driver onl	Rear	No	USA	Buick Roadmaster	
10	Buick	Mastercar	Renee Elis	Israel	Riviera	Midsize	Driver onl	Front	No	USA	Buick Riviera	
11	Cadillac	Visa	Aidan	France	DeVille	Large	Driver onl	Front	No	USA	Cadillac DeVille	
12	Cadillac	Diners	Stacy	United Sta	Seville	Midsize	Driver & P	Front	No	USA	Cadillac Seville	
13	Chevrolet	Amex	Heidi	Netherlan	Cavalier	Compact	None	Front	Yes	USA	Chevrolet Corvette	
14	Chevrolet	Mastercar	Sean	United Sta	Corsica	Compact	Driver onl	Front	Yes	USA	Chrylser Concorde	
15	Chevrolet	Visa	Georgia	United Sta	Camaro	Sporty	Driver & P	Rear	Yes	USA	Chrysler LeBaron	
16	Chevrolet	Visa	Richard	United Sta	Lumina	Midsize	None	Front	No	USA	Chrysler Imperial	
17	Chevrolet	Diners	Leanne	Ireland	Lumina_A	Van	None	Front	No	USA	Acura Integra	

Automobile dataset

Figure 12. Automobile Dataset

Now, we will open this Automobile dataset in the Weka Explorer.

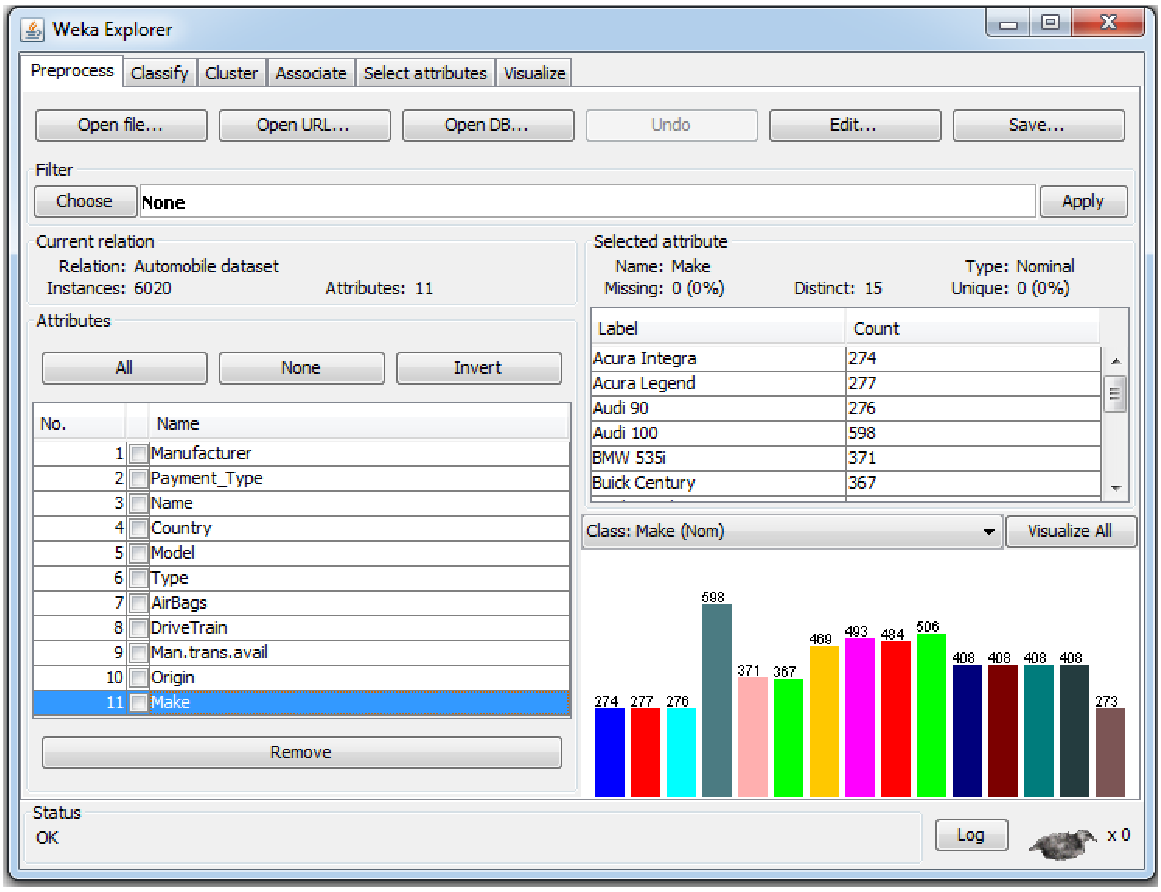

Figure 13. Automobile Dataset opened in Weka Explorer

Figure 13 shows the different attributes of Automobile dataset. It shows that dataset has 11 attributes. In the above figure 13, 15 distinct labels present for the attribute make are depicted.

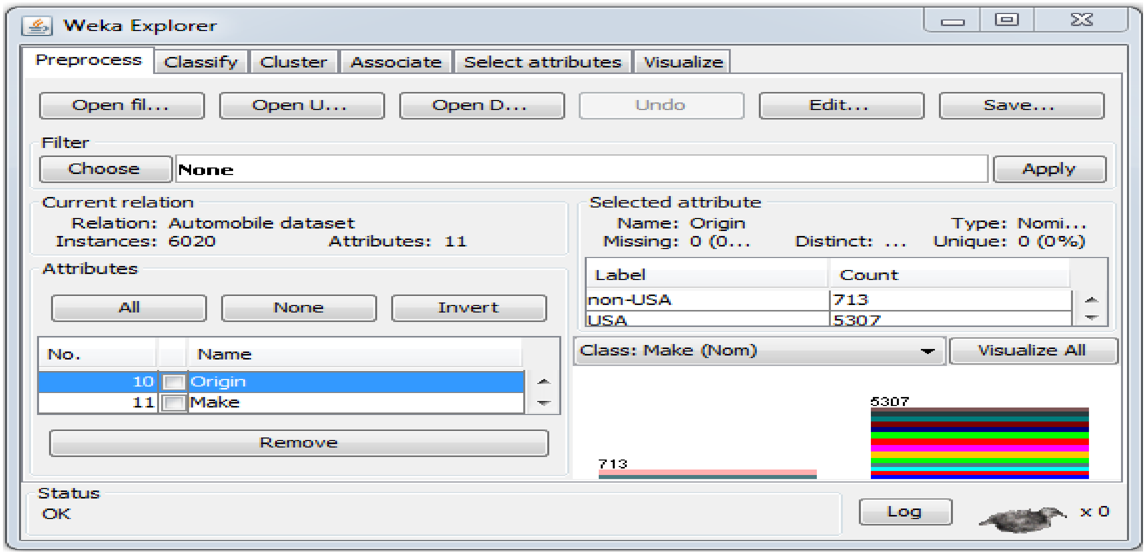

Figure 14. Visualization of Origin Attribute of Automobile dataset

Figure 14 shows that the attribute 'origin' has two distinct labels i.e. USA and non-USA. How different 'make' labels are associated with the 'origin' attribute is depicted in the figure.

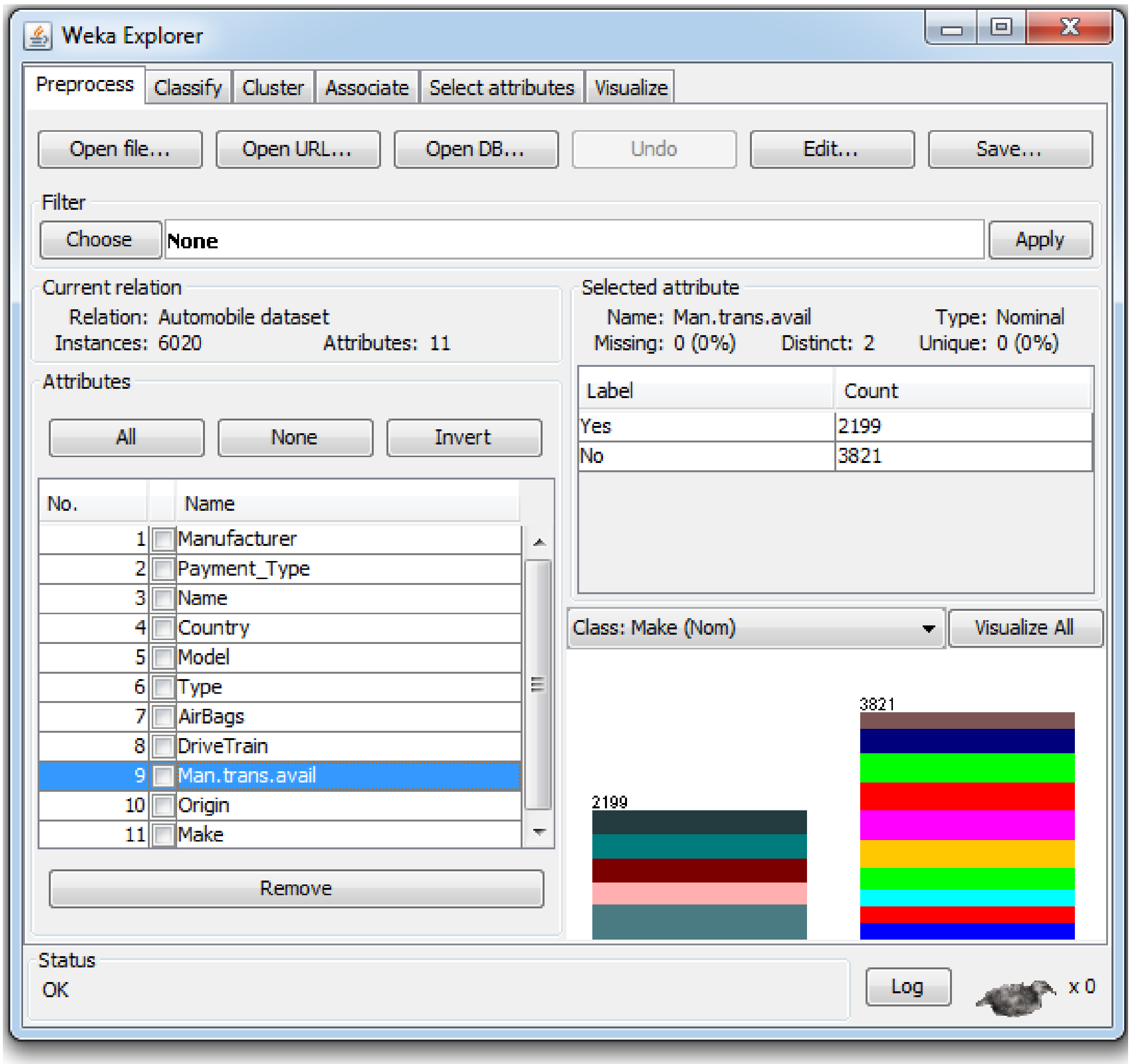

Figure 15. Visualization of Man.Trans.Avail Attribute of Automobile dataset

Figure shows how 'man.Trans.avail' attribute is related to the 'Make' attribute. This attribute has two distinct labels named 'Yes' and 'No'. It shows the values of the 'make' in which manual transmission is available and those in which it is not available.

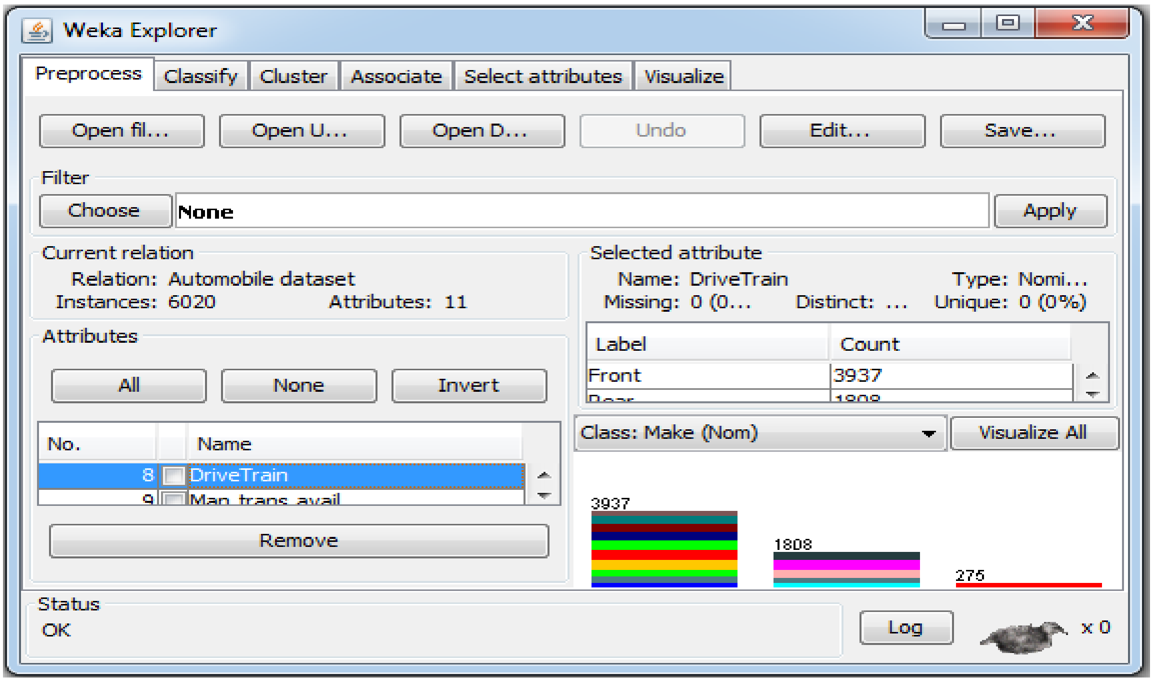

Figure 16. Visualization of DriveTrain Attribute of Automobile dataset

In this figure, it shows that 'Drive Train' attribute have 3 distinct labels. These are: Front, Rear, and 4WD (four wheel drive).

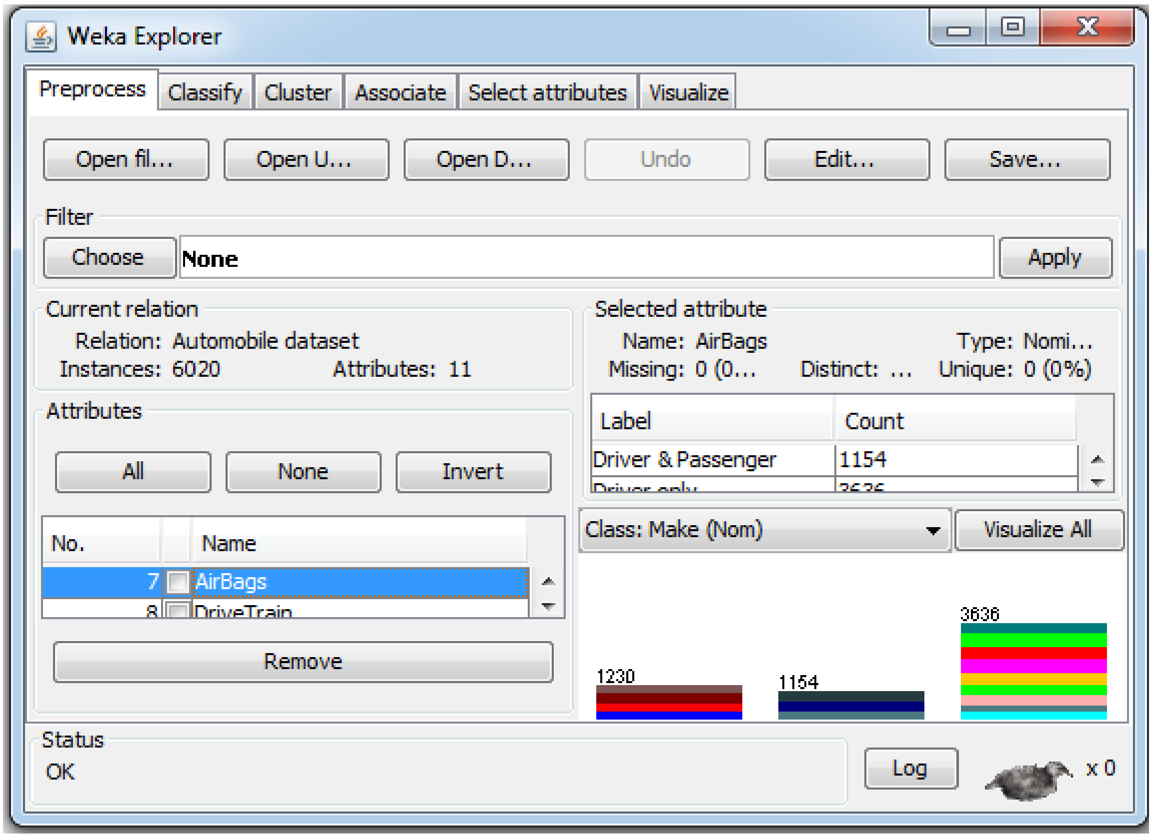

Figure 17. Visualization of AirBags Attribute of Automobile dataset

In Figure 17, Attribute named 'Airbags' has three distinct labels. These are None, Driver and Passenger, Driver only. Figure also shows how this attribute is dependent on 'Make Attribute'.

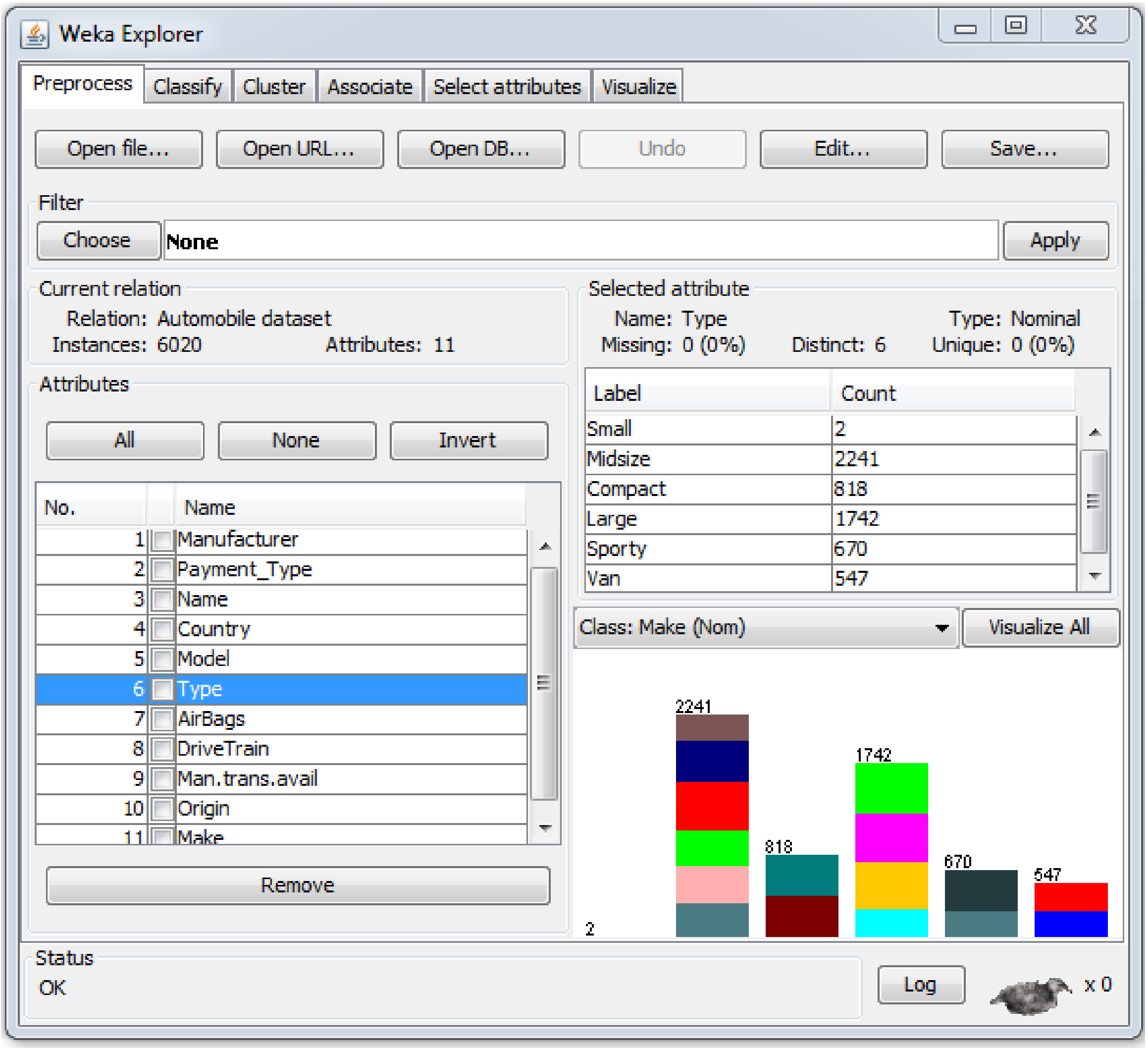

Figure 18. Visualization of Type Attribute of Automobile dataset

Figure 18 shows the distinct labels for the 'Type' attribute. Labels are named as Small, Midsize, Compact, Large, Sporty, Van. It shows the 'type' to which a particular label of the attribute 'Make' belongs to.

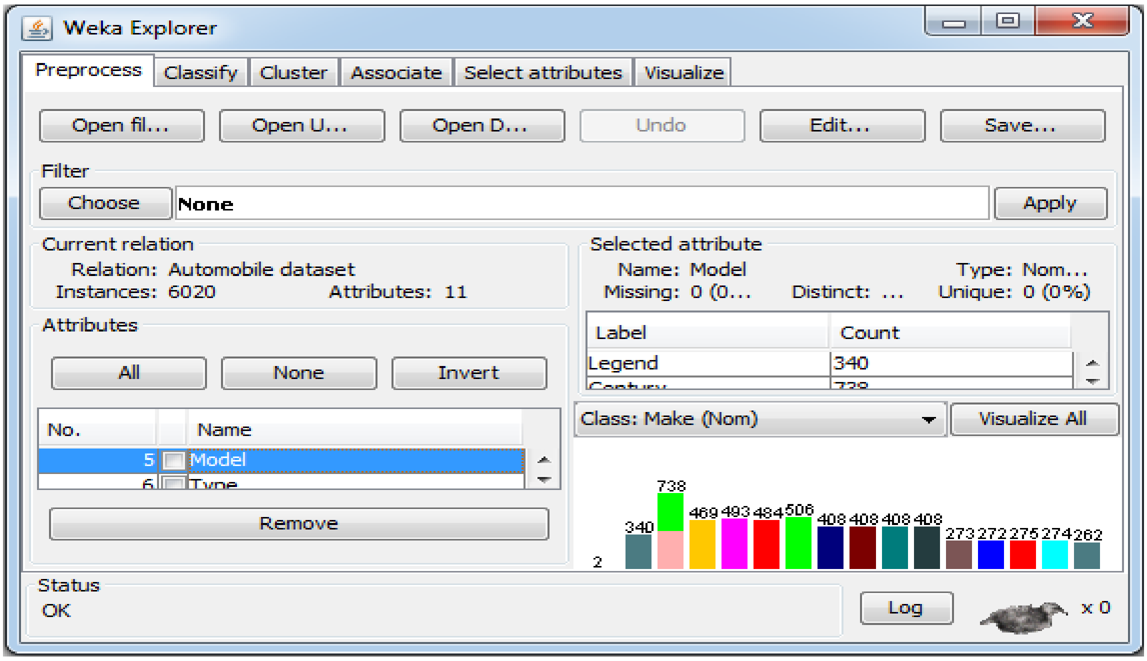

Figure 19. Visualization of Model Attribute of Automobile dataset

In this figure, the characteristics of the attribute ‘model’ is shown. The ‘model’ attribute has 16 distinct labels.

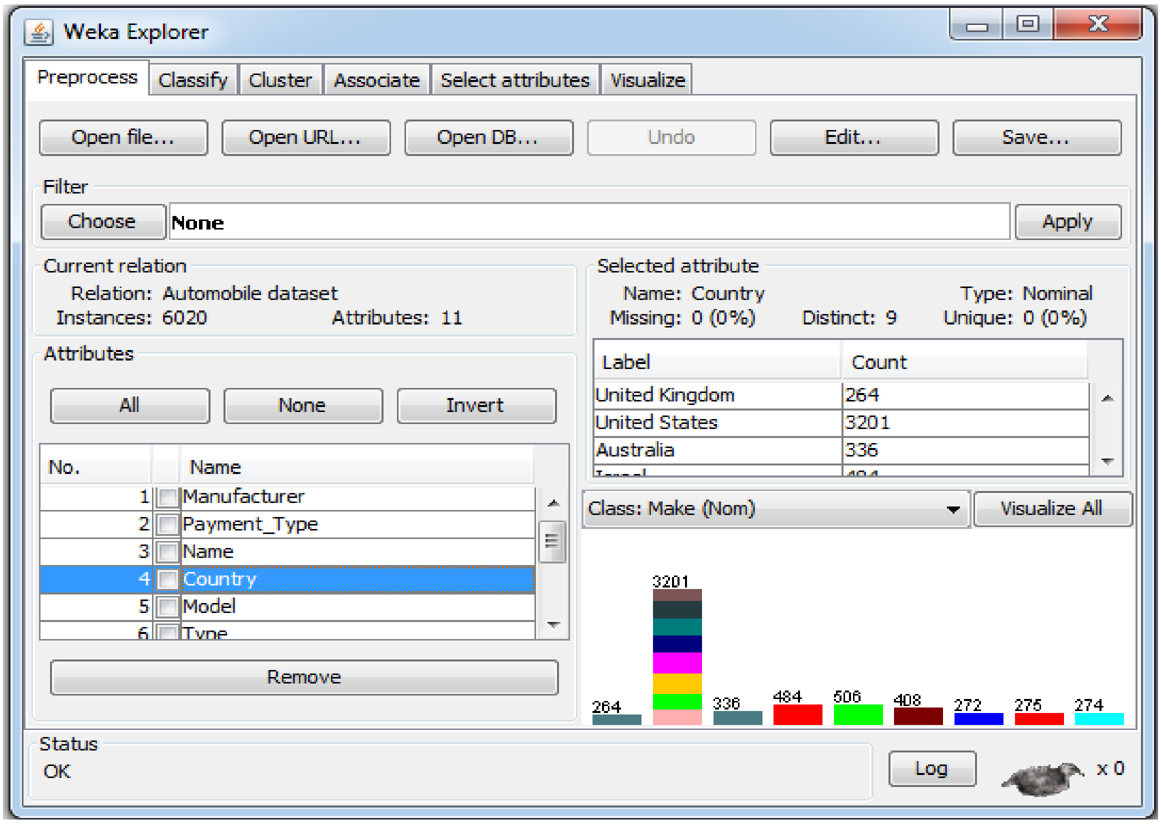

Figure 20. Visualization of Country Attribute of Automobile dataset

In Figure 20, the attribute 'Country' is visualized. This attribute has 9 distinct labels. This attribute is also dependent on the 'Make' attribute and their association is shown in the figure.

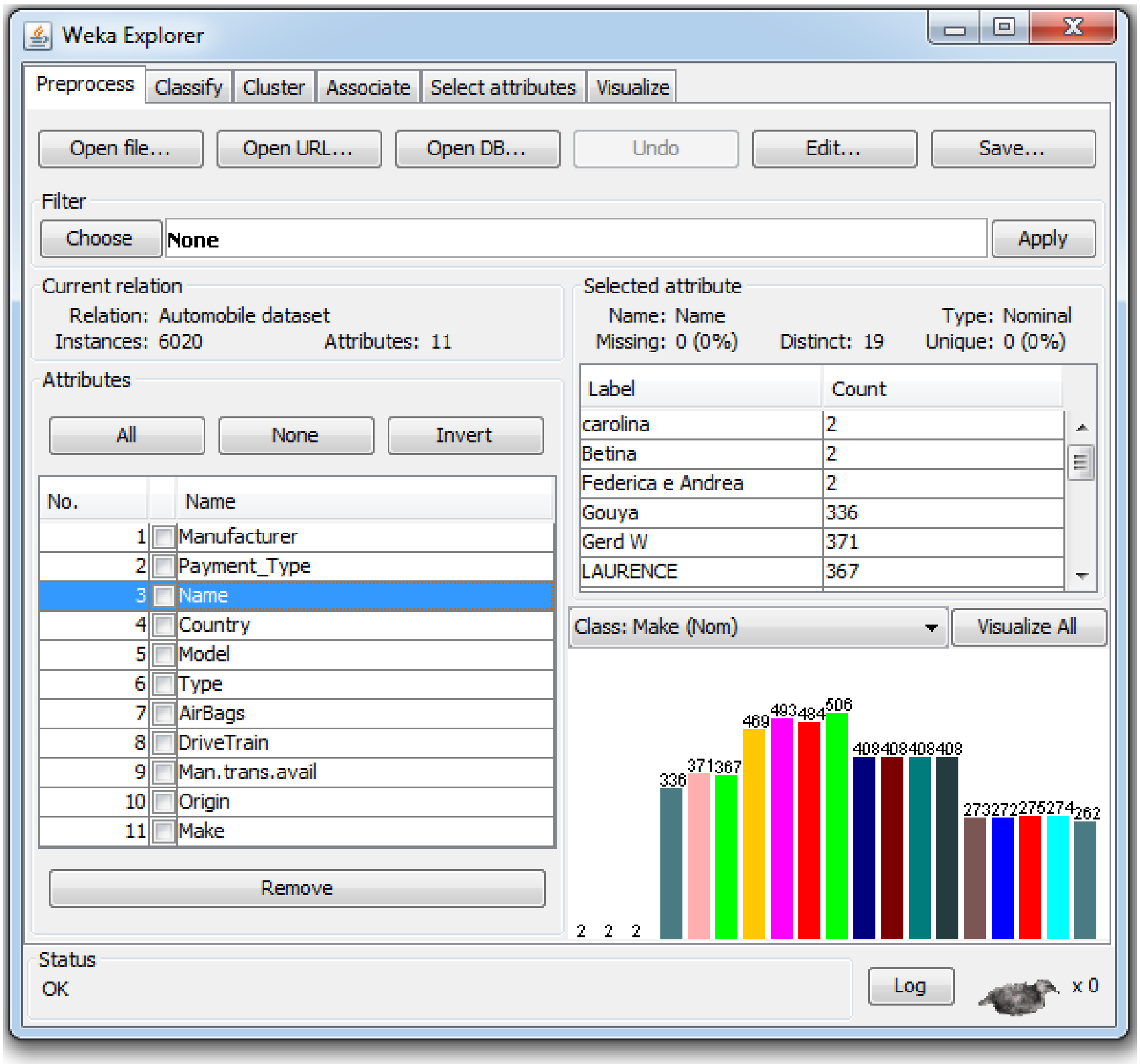

Figure 21. Visualization of Name Attribute of Automobile dataset

This figure shows the 'Name' attribute and its distinct labels. This attribute has 19 distinct values as shown in the figure. The visualization of the relationship between the Name and Make attribute is also shown.

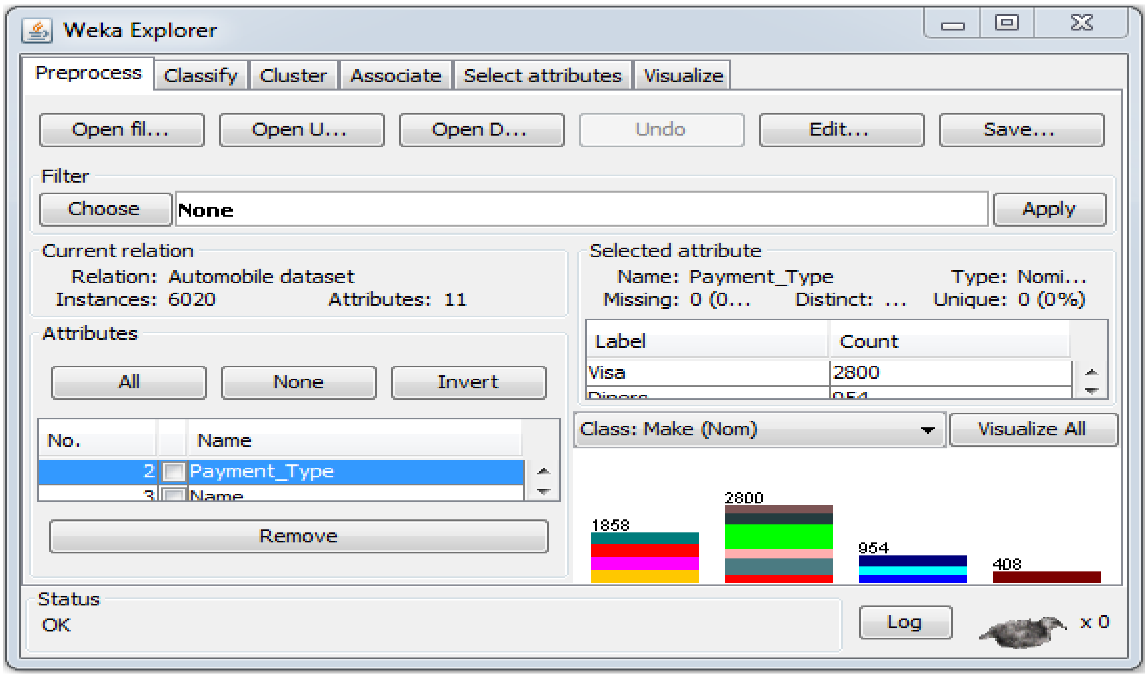

Figure 22. Visualization of Payment Type Attribute of Automobile dataset

Figure shows the visualization of the attribute named 'Payment Type'. It has four distinct labels. These are Mastercard, Visa, Diners, Amex. The figure shows the association between the class attribute i.e. 'Make' and the 'Payment Type'.

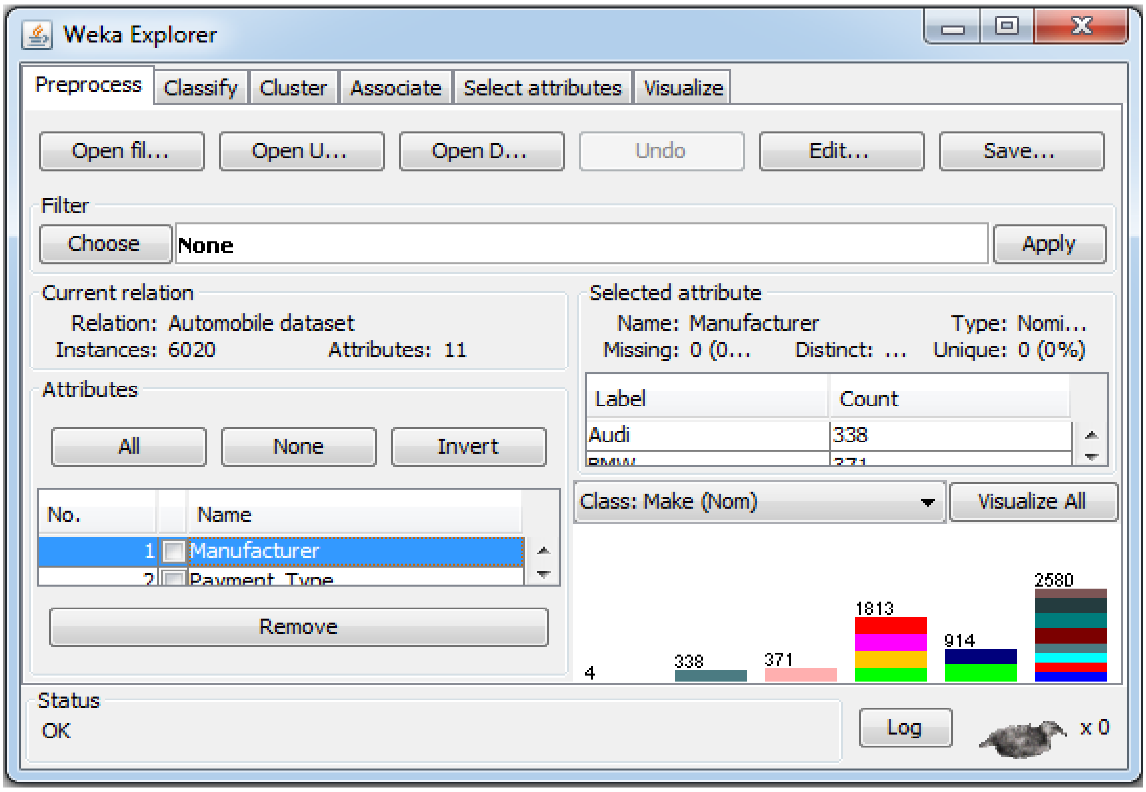

Figure 23. Visualization of Manufacture Attribute of Automobile dataset

Figure 23 shows the attribute 'Manufacturer'. It has six distinct labels. These are Acura, Audi, BMW, Buick, Cadilac, Chevrolet. The association between this attribute and with the class attribute i.e. 'Make' is visualized in the figure.

Now, after selecting the Automobile dataset we will select the Associate Tab in the explorer. Here, we will first select the existing algorithm for Apriori and see the results obtained using it.

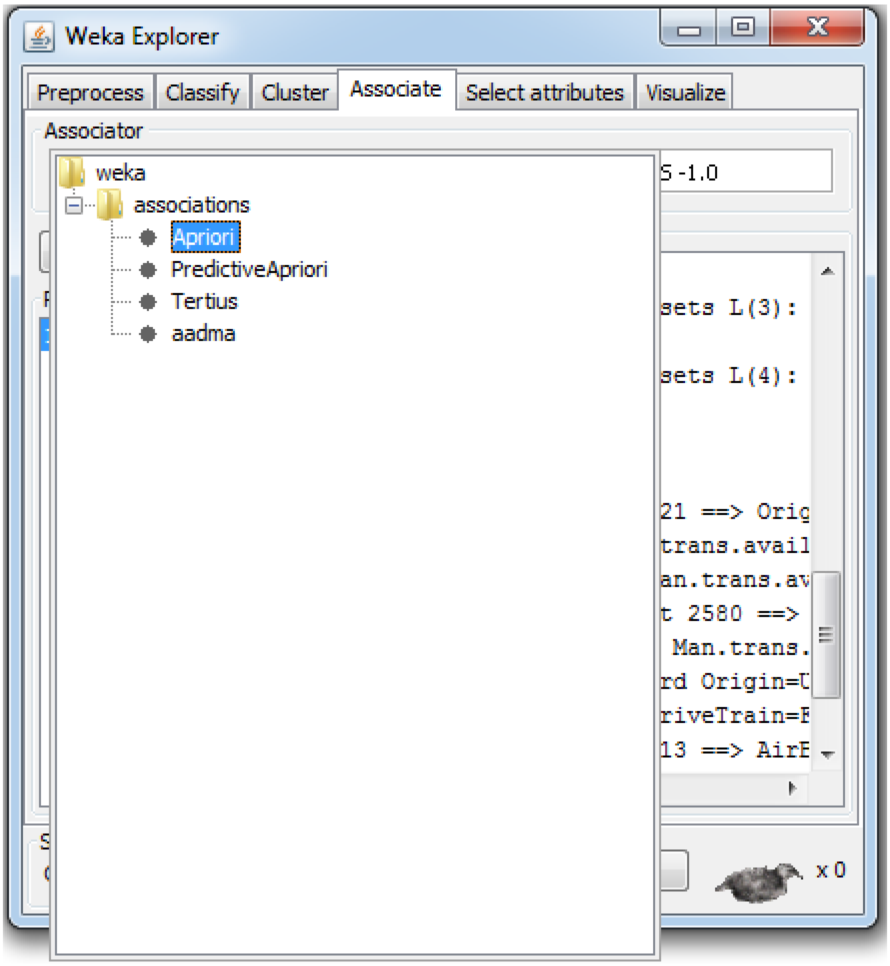

Figure 24. Selection of Apriori algorithm.

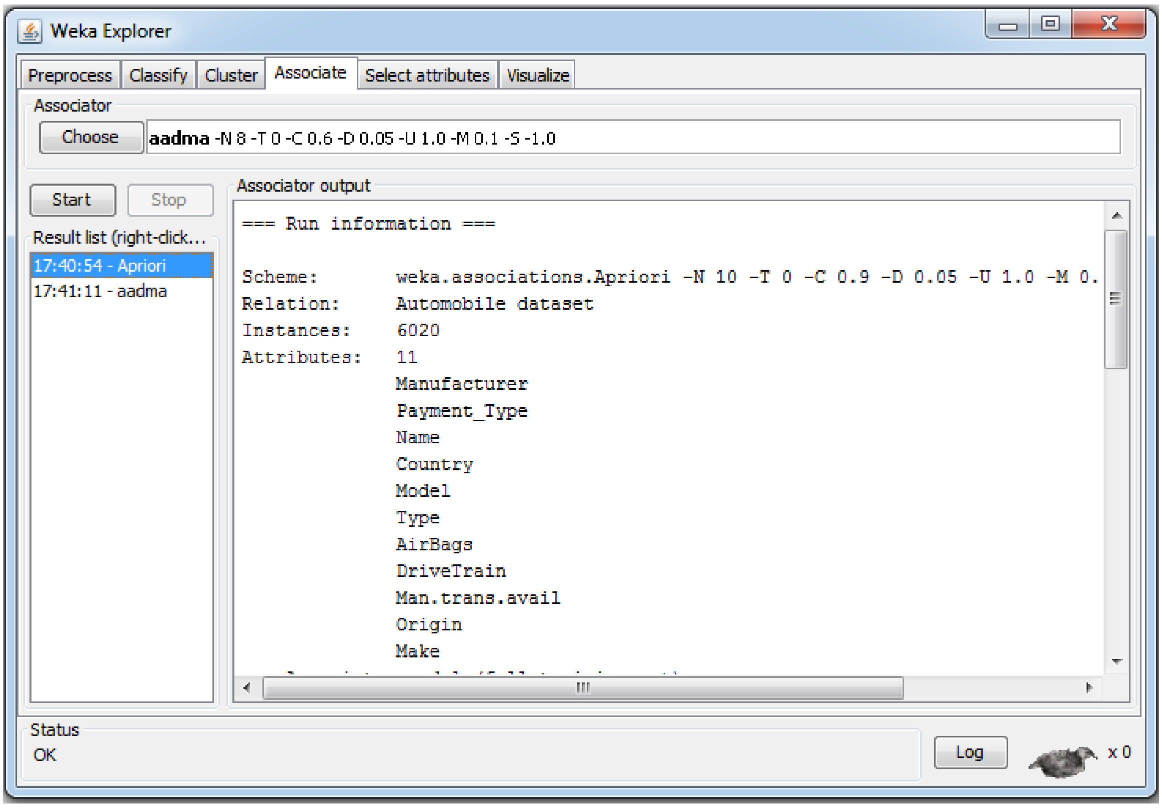

Figure 25. Run information of Apriori

The figure shows the run information of the existing Classical Apriori Algorithm on Automobile dataset. The existing Apriori algorithm is run by using Automobile dataset. The figure displays the number of instances which are covered by the algorithm and the name of the attributes of the dataset. As shown in the figure , 6020 instances and 11 attributes are covered by the algorithm.

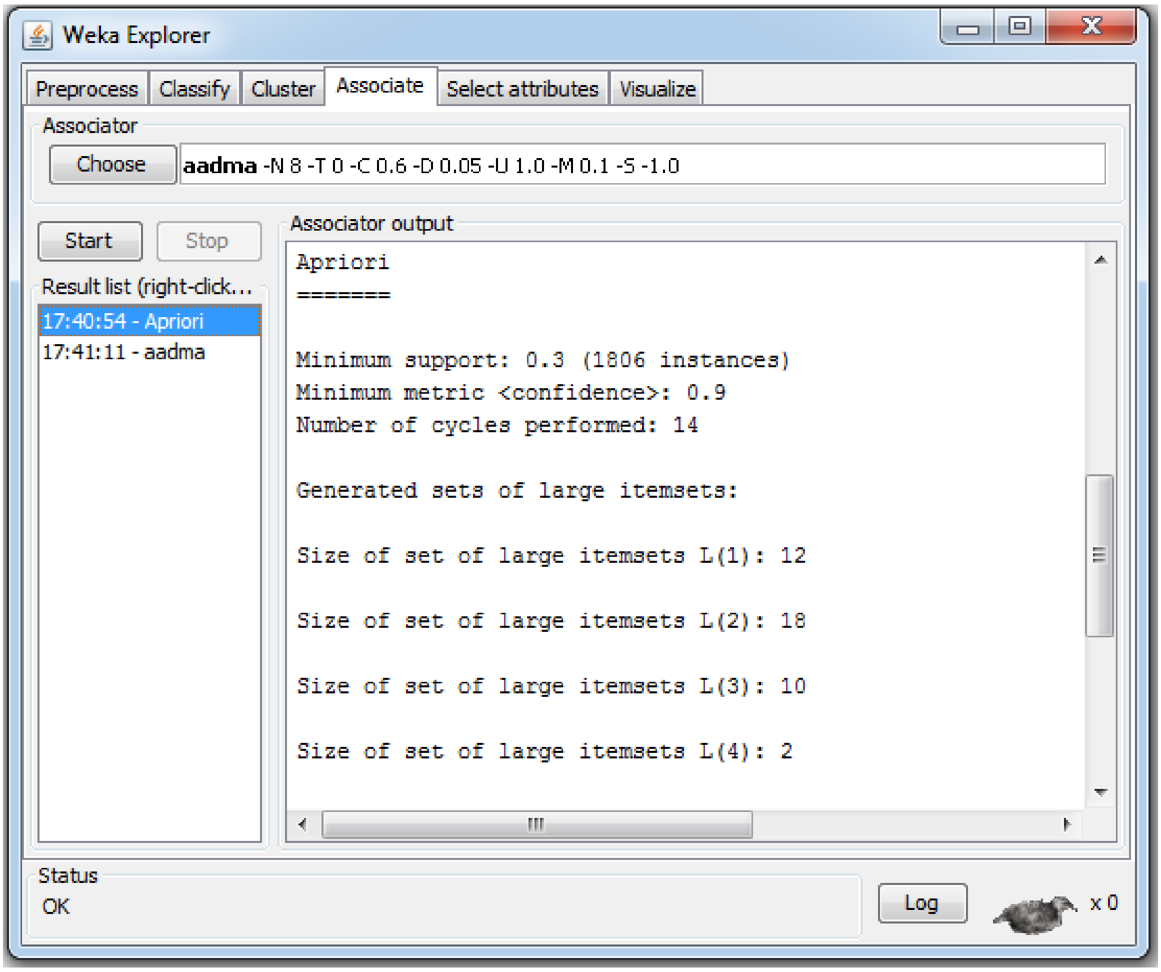

Figure 26. Results of Apriori on Automobile dataset

The above figure displays the results produced by the existing Apriori algorithm on our automobile dataset. It shows the minimum support of 0.3 that is used by it. By using this support, the number of instances that comes under it are 1806. The figure displays the minimum confidence i.e. 0.9 which is used by the algorithm. It displays that the algorithm generated large itemsets i.e. L1 of size 12, L2 of size 18, L3 of size 10 and L4 of size 2. Thus, four large itemsets are produced by the algorithm and the algorithm has taken 14 cycles to do the same. It means the number of iterations are 14.

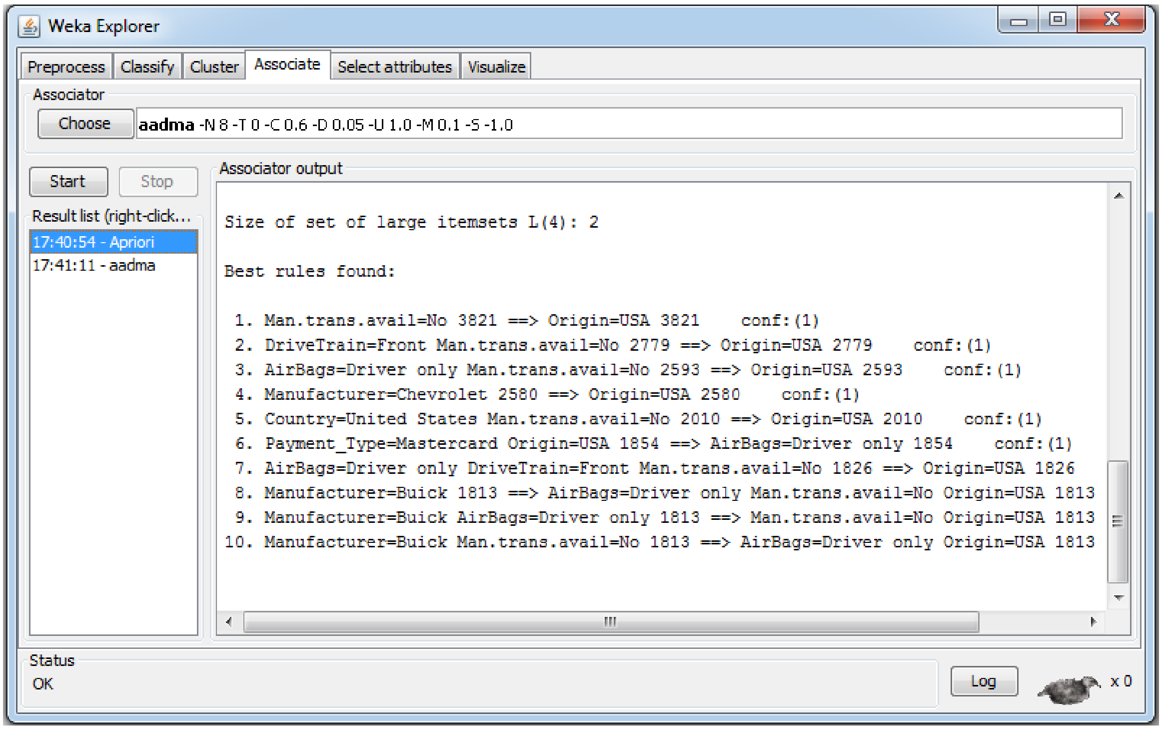

Figure 27. Rules generated by Apriori on Automobile dataset

The above figure displays the Best rules that are generated by the existing Apriori Algorithm. The number of rules produced by the existing algorithm on our automobile dataset are 10.

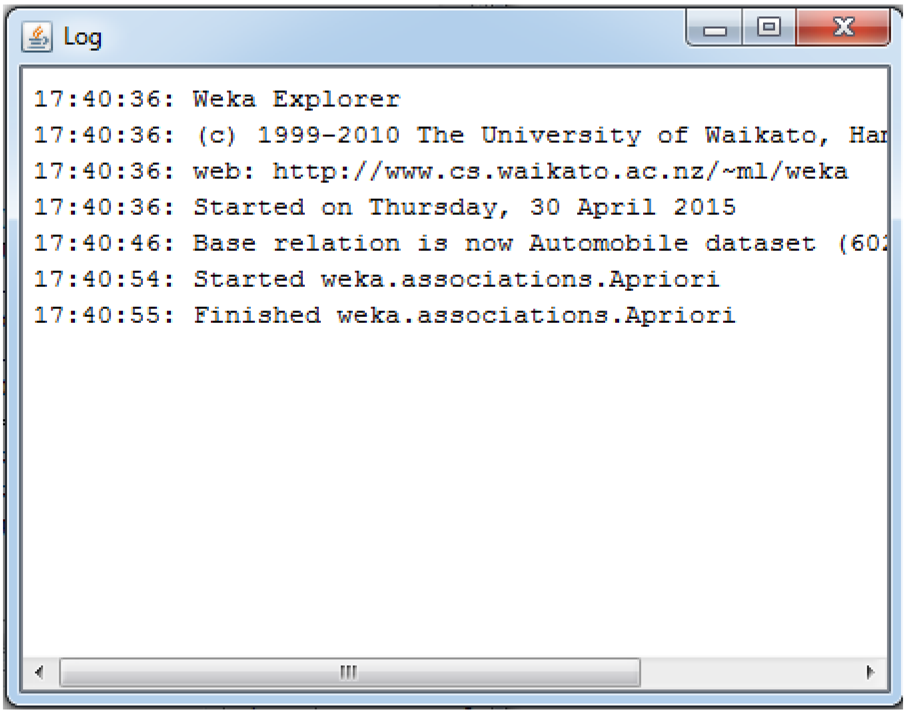

Figure 28. Log information of Apriori on Automobile dataset

Figure 28 shows the time taken by the existing algorithm to perform the operations i.e. to generate the final rules. It is clear from the figure that the algorithm has taken 1 second to run on our automobile dataset.

Now, by clicking on choose, we will select the algorithm developed by us that is AADMA and see the results.

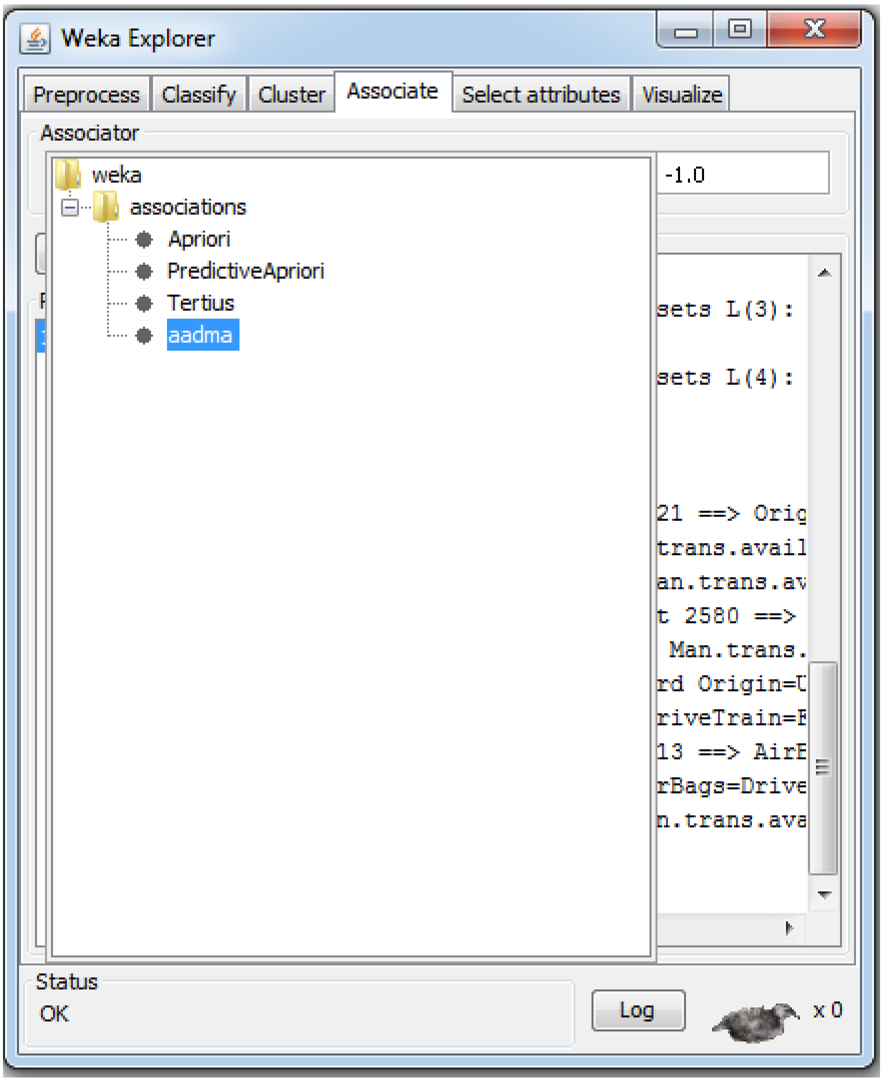

Figure 29. Selection of AADMA

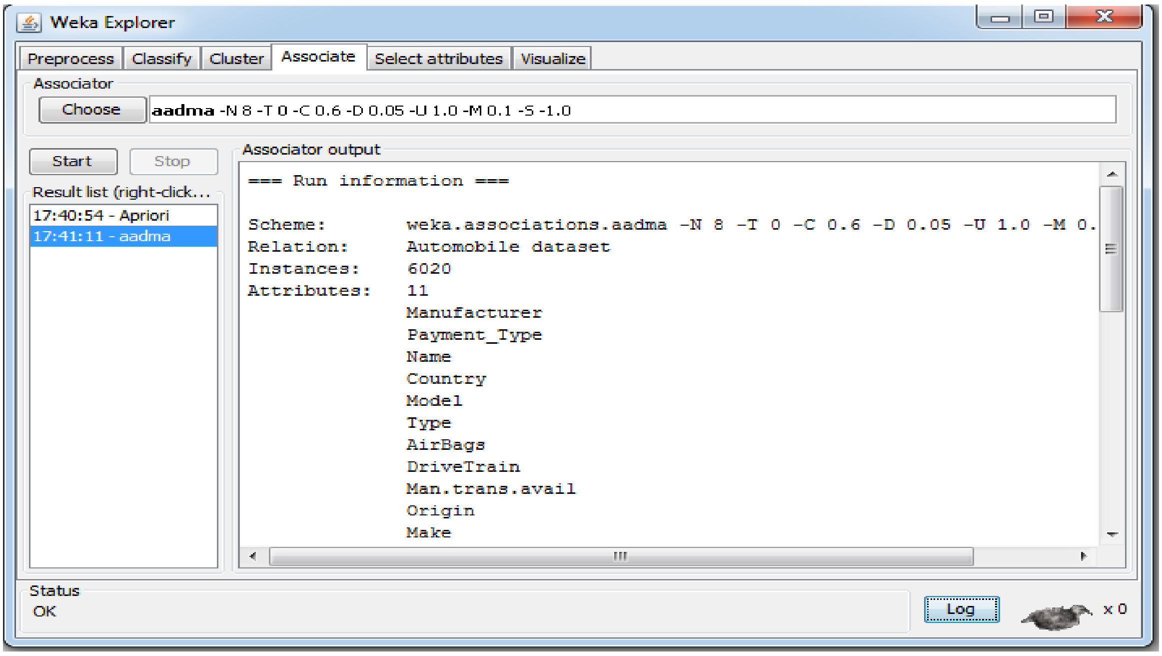

Figure 30. Run information of our AADMA

The above figure shows the run information of our developed AADMA algorithm. It is developed by enhancing the existing Apriori algorithm. AADMA is run by using the Automobile dataset. the figure shows the number of instances covered by the AADMA and the attributes taken into account. As shown in the figure, 6020 instances are covered and the number of attributes are 11.

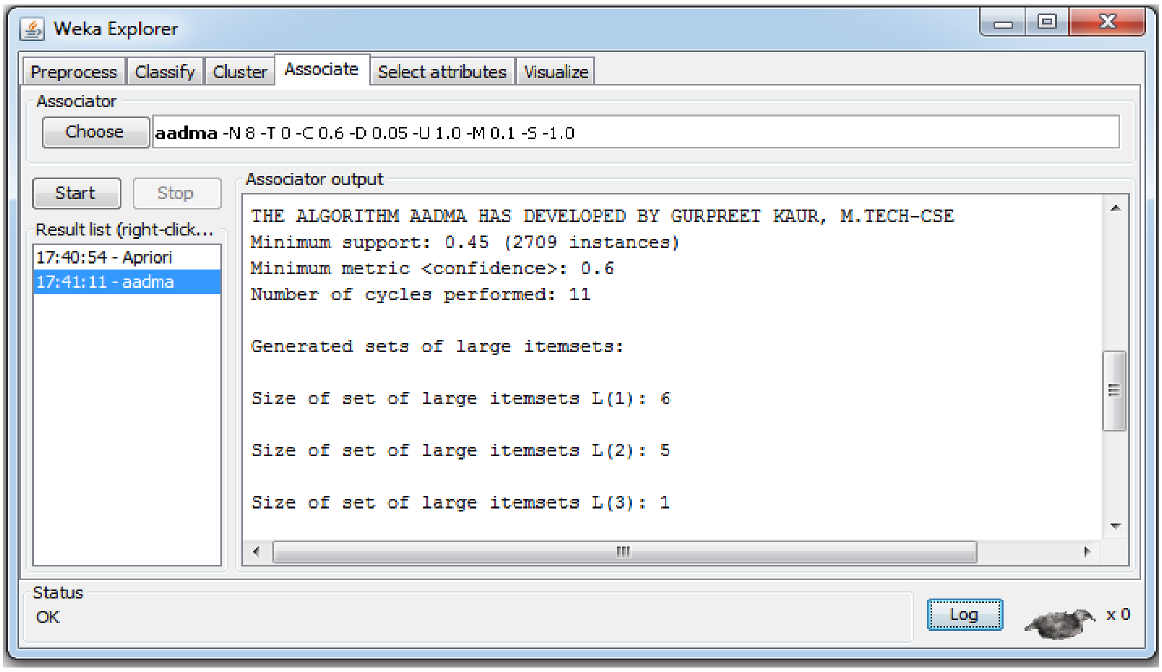

Figure 31. Results of AADMA on Automobile dataset

Figure 31 displays the results produced by the AADMA algorithm on our automobile dataset. It shows the minimum support of 0.45 that is used by it. By using this support, the number of instances that comes under it are 2709. The figure displays the minimum confidence i.e. 0.6 which is used by the algorithm. It displays that the algorithm generated large itemsets i.e. L1 of size 6, L2 of size 5, L3 of size 1. Thus, three large itemsets are produced by the AADMA algorithm and the algorithm has taken 11 cycles to do the same. It means the number of iterations are 11.

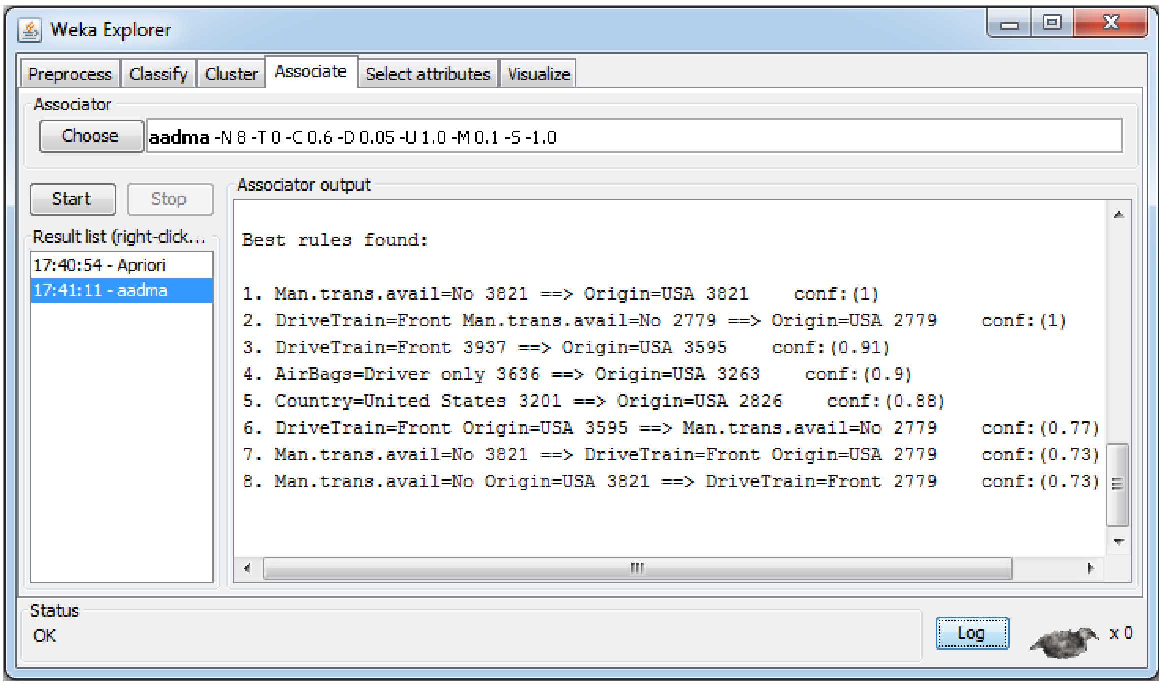

Figure 32. Rules generated by AADMA on Automobile dataset

The above figure displays the Best rules that are generated by the AADMA Algorithm. The number of rules produced by the AADMA algorithm on our automobile dataset are 8.

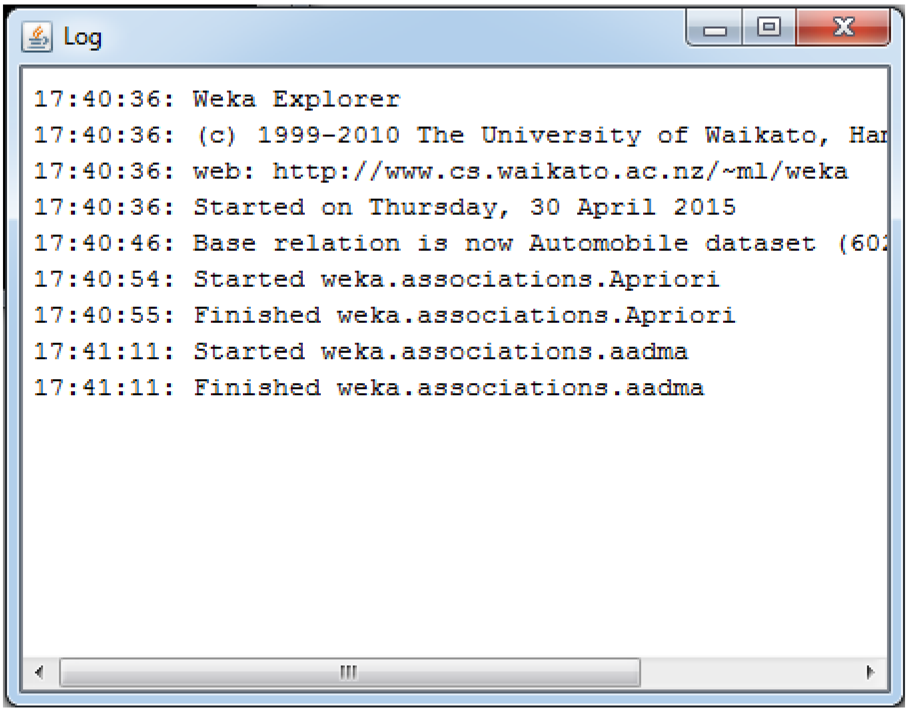

Figure 33. Log information of AADMA on Automobile dataset

Figure 33 shows the time taken by the AADMA algorithm to perform the operations i.e. to generate the final rules from the Automobile dataset. It is clear from the figure that the algorithm has taken 0 second to run on our automobile dataset.

Description	Apriori	AADMA
Minimum Support	0.3	0.4

Table 2. Minimum support used by Apriori and AADMA

The above table displays the minimum support 0.3 that is used by Apriori algorithm. The minimum support used by AADMA is 0.4. Thus, on the same Automobile dataset when we apply the AADMA then more data is covered and the chances of the important rules to be left hidden are very much reduced.

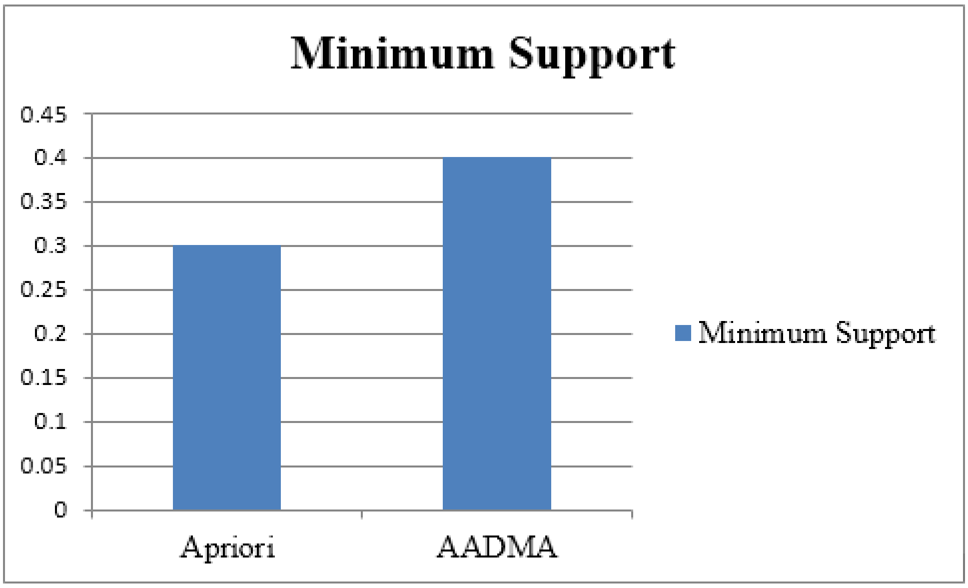

Figure 34. Minimum support used by Apriori and AADMA

Graph depicting the minimum support used by the Apriori and AADMA algorithm.

Description	Apriori	AADMA
Minimum Confidence	0.9	0.6

Table 3. Minimum Confidence used by Apriori and AADMA

The above table displays the minimum confidence that is used by the Apriori algorithm and by the AADMA algorithm. As is clear, on the same dataset of automobiles, the Apriori produced rules by using the confidence of 0.9 while AADMA produced the rules by using confidence parameter with value 0.6.

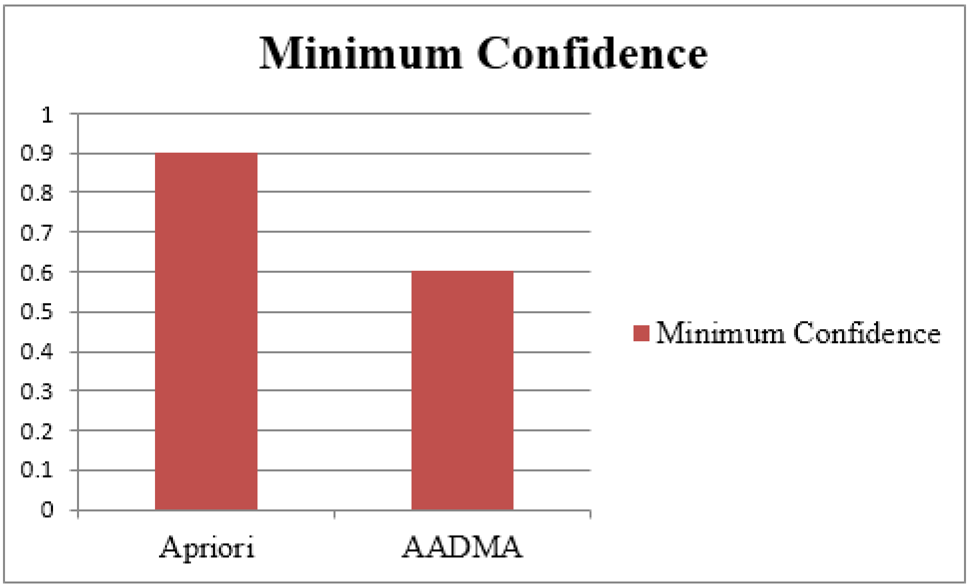

Figure 35. Minimum confidence used by Apriori and AADMA

Graph depicting the minimum confidence used by the Apriori and AADMA algorithm.

Description	Apriori	AADMA
Number of Cycle Performed	14	11

Table 4. Reduction in Number of cycles by AADMA as compared to Apriori

As it is clear from the table, the number of cycles performed by the Apriori algorithm on the Automobile dataset to generate the large itemsets and the rules are 14. While the our developed algorithm AADMA when applied on the same Automobile dataset produced large itemsets and the rules in just 11 iterations or cycles. Thus, our objective of reducing the iterations is achieved.

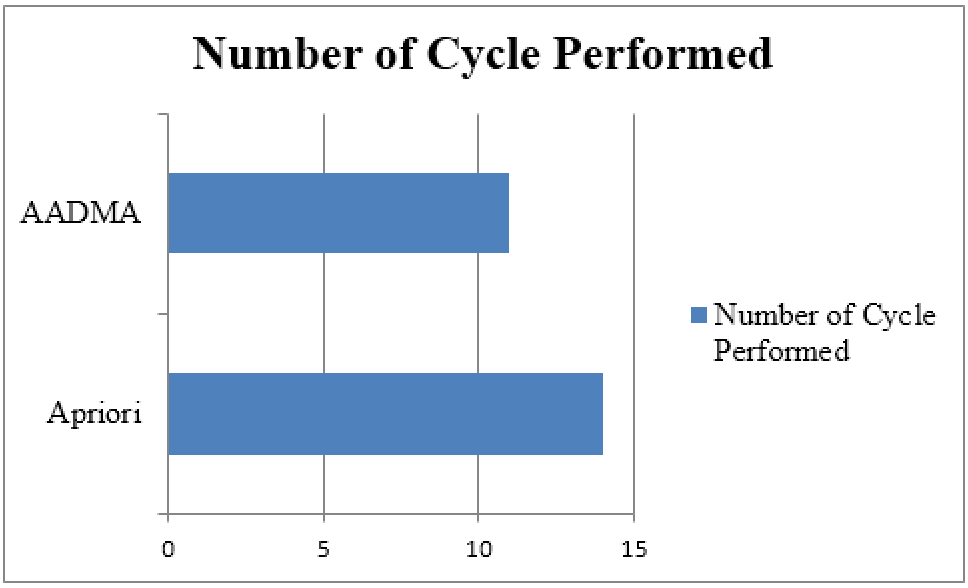

Figure 36. Reduction in Number of cycle performed by AADMA as compared to Apriori

Graph depicting the number of cycles performed by the Apriori and AADMA algorithm.

Description	Apriori	AADMA
Set of Large Itemsets	4	3

Table 5. Reduction in set of Large Itemsets by AADMA as compared to Apriori

Table 5 shows that when our AADMA algorithm is applied on the Automobile Dataset then the total number of Large itemsets produced are 3. But when the Apriori algorithm was applied on the same Automobile dataset then the number of Large Itemsets produced was 4. Thus, our objective of reducing the set of Large Itemsets is achieved.

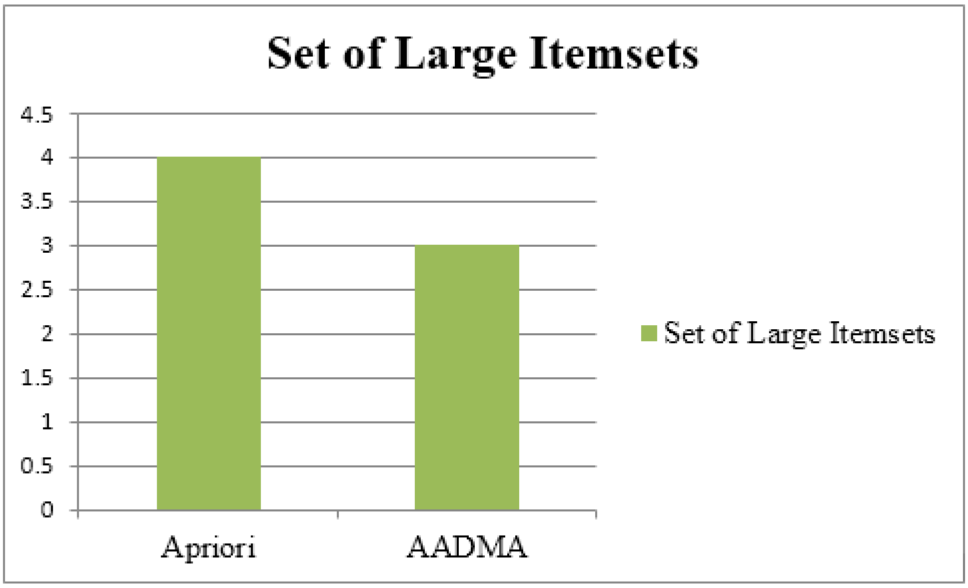

Figure 37. Reduction in set of Large Itemsets by AADMA as compared to Apriori

Graph depicting the Set of large itemsets produced by Apriori and AADMA Algorithm.

Description	Apriori	AADMA
Best Rules	10	8

Table 6. Best Rules generated by Apriori and AADMA

The above table displays the Number of rules generated by the Apriori and AADMA Algorithm. The number of rules generated from Automobile Dataset when Apriori is applied is 10. While, when AADMA is applied on the same dataset of the Automobiles then the rules generated are 8. Thus, our objective of reducing the number of rules while covering more data is achieved.

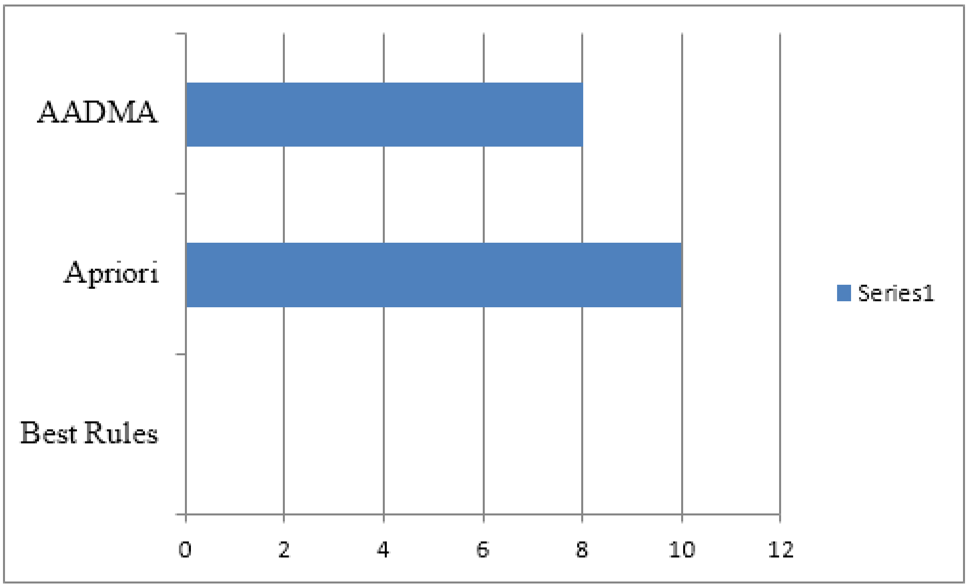

Figure 38. Best Rules generated by Apriori and AADMA

Graph depicting the Best rules generated by the Apriori and the AADMA algorithm.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

Mining data in every domain of research has become a very big issue. Biggest tasks are mining data with very much accuracy and at the same time consuming as less time as possible. In our research work, we have enhanced the existing algorithm named Apriori which is used for ARM. An algorithm named AADMA has been developed as an enhancement of the existing Apriori algorithm. Our goal was the optimization of the most time consuming phases of the Apriori algorithm that is the initial iterations during which small itemsets are counted. The main enhancement regard the use of an innovative method for storing candidate itemsets and counting their support. Thus, the number of iterations have been reduced to 11 using AADMA which earlier comes out to be 14 with Apriori on same Automobile Dataset. Size of large itemsets have been reduced to 3 using AADMA which earlier was 4 with Apriori. Using this, the number of rules obtained have been reduced to 8 which earlier were 10 and more data have been covered. With fast processing time, the time complexity has been reduced. To implement my theoretical idea into realization and to see the results analytical tool weka has been used. In future, work can be done to reduce the time complexity of the algorithm.

REFERENCES

[1] Chanchal Yadav, Shuliang Wang, Manoj Kumar (2013) " An Approach to Improve Apriori Algorithm Based On Association rule Mining ", Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT),IEEE, Tiruchengode, July, p.1.

[2] Dong Gyu Lee, Kwang Sun Ryu , Mohamed Bashir ,Jang-Whan Bae , Keun Ho Ryu (2013) "Discovering medical knowledge using association rule mining in young adults with acute myocardial infarction", Journal of medical systems, New York, January, p.1.

[3] Farah Hanna AL-Zawaidah, Yosef Hasan Jbara, Marwan AL-Abed Abu-Zanona (2011) "An Improved Algorithm for Mining Association Rules in Large Databases", World of Computer Science and Information Technology Journal (WCSIT), 2011, p.311.

[4] Gitanjali J, C.Ranichandra, M. Pounambal, (2014) "APRIORI algorithm based medical data mining for frequent disease identification." IPASJ International Journal of Information Technology (IIJIT), April, p.1.

[5] Gurpreet Batra, Alpana Vijay Rajoriya (2014) "An Enhancement of Association Rule Mining Algorithm", International Journal of Engineering and Advanced Technology (IJEAT), p.115.

[6] Jiawei Han, MichelineKamber.(2014) Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, USA.

[7] Jun Yang, Zhonghua Li, Wei Xiang, Luxin Xiao (2013) "An Improved Apriori Algorithm Based on Features", 9th International Conference on Computational Intelligence and Security, IEEE, 2013, p.125.

[8] Marghny H. Mohamed, Mohammed M. Darwieesh (2014) "Efficient mining frequent itemsets algorithms", International Journal of Machine Learning and Cybernetics, Berlin, April, p.823.

[9] Mohammed Al-Maolegi, Bassam Arkok (2014) "An Improved Apriori Algorithm for Association Rules", International Journal on Natural Language Computing (IJNLC), February, p.21.

[10] Raval Kalyani M, (2012) "Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering ,October , p.439.

[11] Rupinder Kaur, Rajeev kumar Bedi, and Sunil Kumar Gupta, (2014) "Review of association rule mining", International Journal of Advanced Technology & Engineering Research (IJATER), March, p.14.

[12] U.C. Moharana, and S. P. Sarmah, (2014) "Determination of optimal kit for spare parts using association rule mining", International Journal of System Assurance Engineering and Management Springer, Sweden, February, p.1.

[13] Yanxi Liu, (2010) "Study on application of apriori algorithm in data mining", Second International Conference on Computer Modeling and Simulation, IEEE, 2010.

APPENDIX

ARM	Association Rule Mining
AADMA	Advanced Association Data Mining Algorithm