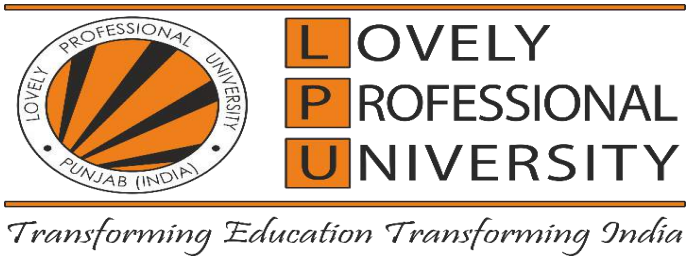

"A New Algorithm for Frequent Itemsets Generations"

A Dissertation Report

Submitted by

MANJIT KAUR

(11306800)

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. URVASHI GARG (16983)

Asst. Professor, LPU.

(MAY 2015)

PAC APPROVAL FORM

LOVELY PROFESSIONAL UNIVERSITY
Transforming Education, Transforming India

School of:Computer Science and Engineering

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the student	:Manjit Kaur	Registration No	:11306800
Batch	:2013-2015	Roll No	:RK2306A23
Session	:2014-2015	Parent Section	:K2306

Details of Supervisor:

Name	:Urvashi Garg	Designation	: Assistant Professor
UID	:16983	Qualification	: M.Tech
		Research Exp.	:4 year

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

Proposed Topics:-

A new
1. Data Mining algorithm for frequent Itemset calculation.
2. Algorithm for cluster formation.
3. Retina scan recognition.

Urvashi 16/9/15
Signature of supervisor

PAC Remarks:

'Topic 2' is approved.

'Publication is expected'

APPROVAL OF PAC CHAIRMAN Signature: 16/9/15 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

Frequent itemset calculation plays crucial role in most of the organization. Frequent itemsets are those items which are frequently occurring in the database. A Lot of algorithms have already been designed for generating frequent itemsets. One of the algorithms that we have used is Eclat algorithm. Eclat algorithm uses vertical dataset and bottom up approach for searching items in database. But eclat algorithm has some limitations. For example, large number of iterations is required for processing the items and more escape time is required for finding frequent itemsets. Eclat algorithm uses bottom up approach that is very complex. We have improved eclat algorithm to reduce escape time and number of iterations by using top down approach and transposing the original database of proposed algorithm

Keywords-Frequent Itemset Mining, Eclat Algorithm, Advanced eclat algorithm, Transposed Database Algorithm.

CERTIFICATE

This is to certify that **MANJIT KAUR** has completed M. Tech dissertation titled **"A New Algorithm for Frequent Itemsets Generations"** 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 proposal is fit for the submission and the partial fulfilment of the conditions for the award of M. Tech, Computer Science and Engineering.

Date: **Signature of advisor**

ACKNOWLEDGEMENT

I am highly thankful to my mentor Urvashi Garg who gives me proper guidance for my dissertation. With her inspiration & suggestion I am being able to do this research. Under her supervision I am going to do "A New Algorithm for Frequent Itemsets Generations" It would really hard to make this dissertation without her supervision. I would really express my deeper thankful to her. She plays major role for my dissertation. She always inspired us to do something better.

I also thank to my HOD who provided facilities to me being required.

MANJIT KAUR

DECLARATION

I hereby declare that the dissertation proposal entitled, **"A New Algorithm for Frequent Itemsets Generations"** submitted for M-Tech degree is entirely my original work and all ideas and references have been duly acknowledged .It does not contain any work for award of any other degree or diploma.

Date _____

MANJIT KAUR

Regd.no:11306800

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Data mining is the process of extracting useful information from various sources. These sources may be any relational database, transactional database, data warehouse etc.
Data mining analyzes data from different perspective and summarized all the data into useful information. Data mining is commonly used by various organizations. Like financial, communication organizations. It helps to these organizations for decision making and also determines effect of sales, customer satisfaction and profits. The main goal of data mining is to access valuable information and convert this valuable information into the understandable structure.

1.1 The Foundations of Data Mining

Data mining techniques are defined with the development of research, make some enhancements in data access and new developed Technologies give permissions to users to navigate for using the data. Data mining uses this process for information navigation to users, Data mining can be used for applications. It can be used in the business community. Three technologies are defined for data mining. These technologies are:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases increase at unprecedented rates. From META Group survey of data warehouse projects it was discovered that 19% people are those people who comes under 50 gigabyte levels. Values of these numbers will be more in some industries. Industries may be retail and so on. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms includes various techniques which are defined from last 10 years. Only new techniques are implemented by different method. These methods must be mature, useful, and reliable and can be easily understand by people.

Business data is converted into business information. In this new step is used based on previous step, For example, In data navigation applications, dynamic data access is used

for drill-through and Huge amount of information is used in large databases that is major issues for data mining.

The most commonly used techniques in data mining are:

- **Artificial neural networks**: These networks are non-linear predictive models which can be understood by user through proper training and structure.

- **Decision trees**: Decision trees are those trees which provide decision. This type of tree creates associations' rules for market basket dataset. There are some decision tree methods available. These methods are Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

- **Genetic algorithms**: Optimization techniques are used according to evolution.

- **Nearest neighbour method**: It is one of methods which divide every record. It is based on those classes of records which are similar in old dataset. All records are classified in the dataset.
- **Rule induction**: It contains if then rules. If then rules are those rules in which we check that if condition is satisfied or not. If it is satisfied then only then part is working. For example if there is raining then take umbrella with us.

Data mining is also known as knowledge discovery. We can find the frequent item set by association rules. Associations rules plays important role in different companies, organizations, online markets and network services and so on.

Steps of knowledge discovery in data

1. Data Cleaning: There may be irrelevant data and noise data in the database. By data cleaning step, irrelevant data and noise data are removed from the database. Only clean data is stored in the database.
2. Data Integration: Data from different sources collected together in the one place.
3. Data Selection: Appropriate data is selected from database.
4. Data Transformation: Convert all the data and information into accurate form that is used by data mining.

5. Data Mining: It is very crucial step which includes different techniques. These techniques are basically helps for getting the pattern from database.
6. Evaluation of Pattern: Interesting patterns represent acquaintances. Interesting patterns are included which are based on measures parameters.
7. Knowledge Representation: Knowledge Representation is the last step which is very crucial to provides knowledge to all users in graphically way or any ways which should be understandable by user. We use different visualization techniques.

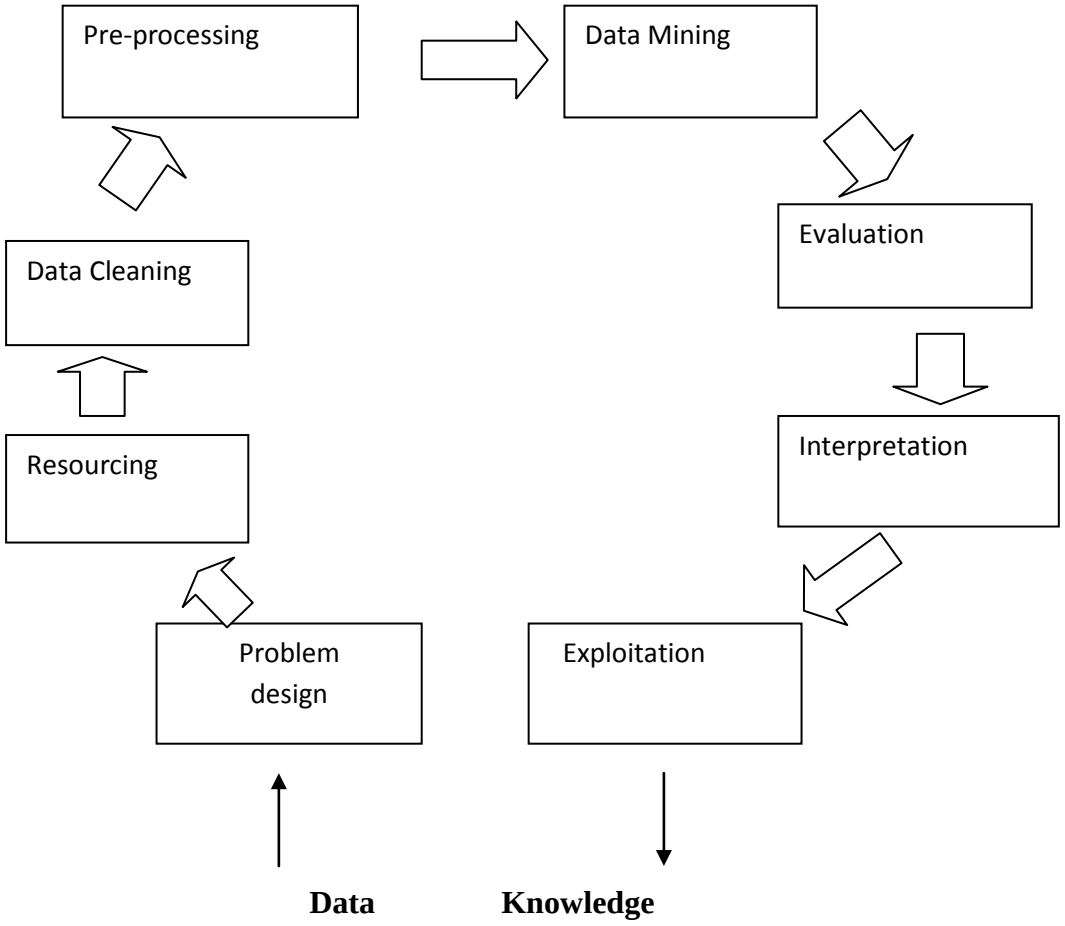

Fig.1.2 Data Mining Process

There are different Data mining functionalities which are described given below:-

1. **Discrimination and Characterization:-**Data characterization provides summary of data which is present in class. And the purpose of data discrimination is to make

comparison between the goal class and one or group of comparative classes. Class/concept descriptions are derived by using both of these functionalities

2. **Cluster Analysis:**-Cluster analysis or clustering is that process which makes group of set of objects. If any object is same in one group then that object must be similar in other group also. By which, cluster analysis can be possible.
3. **Classification and Regression:** - Data mining includes different techniques. Such as Classification, Regression etc. Classification is technique of data mining that plays important role for membership for groups. Prediction of numeric can be possible through Regression analysis.
4. **Outlier Analysis:** - If any object does not exist in object general model in that case those objects comes under outlier analysis. These types of objects are known as outliers and outlier analysis is process of analyzing outlier objects.
5. **The Mining of Frequent Patterns, Associations and Correlations:**-Frequent patterns are those patterns which are occurred frequently in data. Association rule mining is one of method through which interesting frequent patterns or associations are discovered. These interesting patterns are found from transaction databases, relational databases or other databases.

There are huge amount of data generated in the various organizations. Therefore organizer has to take number of decisions during extraction of useful data from the huge amount of data. But it is difficult to extract each and every record. So organizer finds frequently occurring data in the database. Based on those frequent data, organizers takes decision and makes business strategy. Frequent items are also very essential part in the online markets. In online markets wide variety of products are available for customers. But customer gets confused while choosing among those products. So the products which are mostly purchased by various customers (frequently purchased products) help new customer for best product selection. Different algorithms have been developed for calculation of frequent itemsets. Apriori is a very basic algorithm. But it takes a lot of time for calculations of frequent itemsets. In this algorithm we need to calculate support and confidence. We need to scan to database again and again. So FP growth algorithm has been developed. FP growth stands for frequent pattern growth. In this algorithm, database scanning is required two times. Time consumption is more. To remove these limitations, Eclat algorithm was developed. Eclat algorithm is very commonly used algorithm. Eclat algorithm uses vertical database and depth first search approach. All the data is stored in vertical form. Bottom up approach is used for searching items in the

database. Searching starts from bottom to top. Only support is calculated in this algorithm. But it takes more time than top down approach. There is no need to calculate confidence as it is not required and also because confidence increases complexity. Therefore this also not included in proposed algorithm. Firstly we need to calculate support of all items individually. We can decide support in two ways one which is decided by user and second by averaging all items. After calculating support of all items we will compare support of items with minimum decided support. The items which has support more than or equal to decided support are frequent itemsets. This is whole process of eclat algorithm. But this algorithm has limitations. Access time is more in this algorithm. Numbers of iterations are more. Complexity of eclat algorithm is more because of using bottom up approach. This algorithm can be improved by performing transpose operations. In transpose operations, all the items are given to all transactions at same time. It will save time also. Also top down approach is used in proposed algorithm. Number of iteration will be reduced by transposing database.

1.2 Advantages of Data Mining

Marketing / Retail

Data mining plays important role in marketing companies. Market companies develop new models which are based on historical data. It is helpful for predicting to response of new marketing campaigns. These marketing campaigns may be direct mail, online marketing campaign and others. Data mining helps to retailer to select best products for customers. Profits of products also increase by using data mining.

Different benefits are provided to retailer through data mining. Market basket analysis is very beneficial for companies because companies can arrange the products according to frequent purchased products. In this way, Customer can buy products with less time. It is helpful for both retailer and customer. By good arrangement of products, customer takes less time while selection of products. By using market basket analysis, retailer can provide discounts on products by which sale of products will increase.

Finance / Banking

All financial information can be provided by data mining. It provides information about loans. Banks can get idea about loans from historical data. It can also helps for determine fraud people.

Manufacturing

Data gives information about the fault equipments. Through this information, manufacturers can take decision to remove the fault from equipments.

Governments

Data mining plays important role in government agency. It analyzes records of financial transaction which are used to create patterns. These patterns detect information related to criminal.

1.3 Disadvantages of Data Mining

Privacy Issues

There is fear in the mind of people because someone can hack the information of people. Information related to customers can be getting by businesses for knowing buying habits of people. At this time, personal information can be leak.

Security issues

Major issue related to data mining is security. Businesses want all personal information about customers which may be hacked by hackers.

Misuse of information/inaccurate information

Data mining gives information related to customer that can be misuse by unauthorized people.

1.4 Market basket analysis

Market basket analysis is dataset which will be used in research. Market basket analysis is process of finding the buying habits of people. Market basket analysis is very useful term for improving business strategies. Market basket analysis is a data mining techniques for

finding the association between datasets. Huge amount of data stores in the databases. Companies use the association rule mining for finding the associations rules from databases. Associations can be defined in form of association rules. It also helps to define the store layouts. If one item is purchased then it encourage the people to buy other items also. For example if any person buy computer then he will also buy anti-virus for security purpose at same time. So place anti-virus near to computer, so whenever person buys computer then same time they can buy anti-virus. Market basket analysis checks that which item is placed by people in their baskets. By that information businessman, companies and shopkeepers keep their items together and make plans for their shelf space. Market basket analysis tells items that are purchased by seasons.

It helps for promotion of products. We can place the product according to requirements. It provides various promotional opportunities.

Two rules are measures as follow:

- Support: it means that rule X=>Y that is present in the transaction set D1 with support s, where s defines percentage of all transactions in transaction set D1 which includes X U Y. For example, a support of 40% for rule tells that both computer and anti-virus software are purchased together in the store. Both computer and anti-virus buy at same time.
- Confidence: The rule X=>Y that is present in the transaction set D1, where c shows the percentage of transactions in transaction set D1 which are contained X also contain Y. For example, 60% of confidence means that 60% of customers who purchase computers also purchase the anti-virus software.

The items which have support and confidence more than or equal to given minimum support threshold and minimum confidence threshold, are the frequent item sets, these frequent item sets are purchased frequently.

1.5 Data Mining Algorithms for frequent item sets generations

Itemsets are collection of items. Frequent item sets are those items which are frequently occurs in database. A frequent item set mining is one of the data analysis ways that develop for purpose of market basket analysis. Frequent item sets are those items that have support and confidence greater than equal to given minimum support and minimum confidence. It means that it helps the shopkeeper to improve the business strategy to understand which items are frequently demanded by particular customer. It helps in better business environments. For example, when place butter and bread at separate places in

the mall then sale of bread and butter is less. But when butter and bread are put together at same place then sale is increased. Companies try to find association rules. Association rules means find those items which are frequently occurs. There are numbers of algorithm available for finding frequent item sets such as eclat algorithm, apriroi algorithm and so on. The main rule of data mining is to discover all the items that have support and confidence greater than or equal to given minimum support and confidence. It means that it helps the shopkeeper to improve the business strategy to understand which items are frequently demanded by particular customer.

1.5.1 Frequent Item sets and Association Rules

Association rule and frequent item set mining a very popular area for research work. The apriori algorithm includes both frequent item sets and association rules as important terms. Association rules defines items in a given set describes that when item P contain items from database then Q also contains the items of same database.

Using Market Basket Analysis it is analyzed that customer purchase mostly which type of items. Thus frequent item sets are obtained by such analysis using some available techniques.

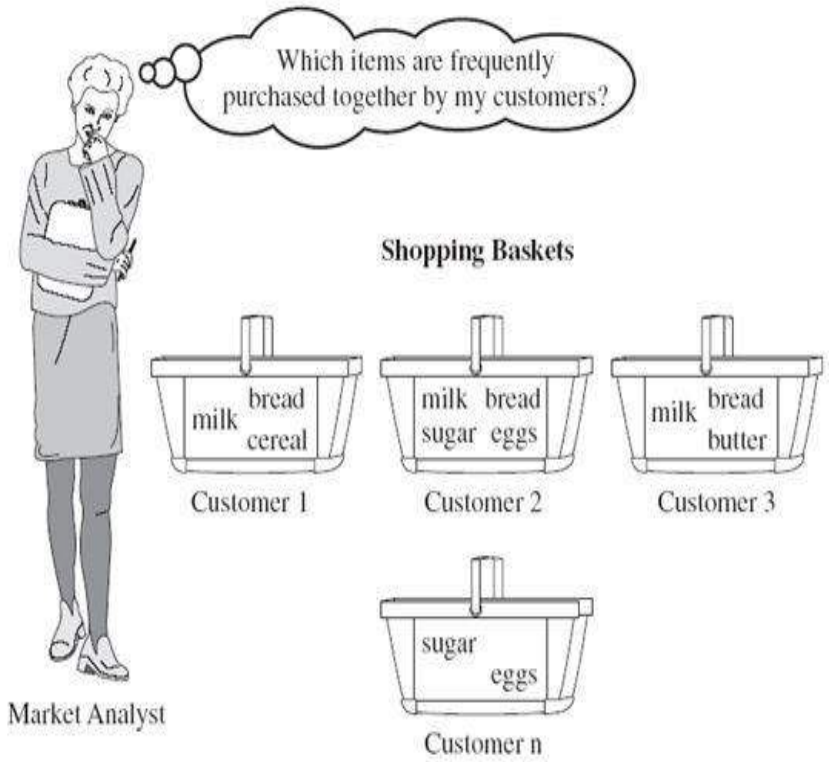

Fig. 1.5 Market Basket Analysis

- Frequent itemsets: which are frequently purchased or in other words those items which satisfying minimum support.
- Association Rules: Frequent item sets are used to create rules.

E.g., Egg → Milk [given support = 5%, confidence = 100%],

A collection of items is known as itemset and an itemset with N number of items is N-itemset.

A frequent item set {E, M} =4

M→E= 100% confidence

1.5.2 Measures in association rules

Two main components are available in association rules.

Support: The support of an item set is the total transactions that contain all the items of that particular item set. That means containing both A and B: A∪B.

Support= P (A∪B).

Confidence: Ratio of support and the support of A is called confidence of any association rule

Confidence= P (*B* | A)

Support and Confidence

Support count: The support count is defined as an item set which contain A item from all transactions. It is represented by A. count.

Support= $((A \cup B).count)/(n)$

Confidence= $((A \cup B).count)/(A.count)$

$$\text{Support (XY)} = \frac{\text{Support count of (X}\cup\text{Y)}}{\text{Total number of transactions in database}}$$

Confidence also includes probability which works with condition that if X occurs in an event then Y will also occur.

$$\text{Confidence (XY)} = \frac{\text{Support count of (X}\cup\text{Y)}}{\text{Support(X)}}$$

The association rule must satisfy the requirement to find the problems within the association that both support and confidence have value not less than the given threshold. If the aim is not taken into consideration that that particular item set is neglected from frequent item set.

1.6 The Apriori Algorithm

The Apriori includes prior knowledge of items which are frequent. This is a step wise search in which by using the previous itemsets, the next level itemsets are found. Apriori algorithm is an advantageous algorithm presented to mine the items which are frequently purchased and also association rules are generated. Firstly, L1frequent -1 itemsets after collecting the occurrence of items by scan the whole data is mined. Then accumulate all the items with minimum support. This result is shown by L1. This frequent-1 itemset required to generate the next level itemsets until it reaches to null value. It needs one full scanning of dataset to find the result of Lk. Apriori property is available in algorithm to reduce the search [1].

Apriori algorithm having mainly two steps which are:

1) Join step: The frequent item sets are joined step wise to find the candidates.

2) Prune step: When the items have support less as compared to the given minimum support those items are deleted under this step. Also, the item set having no sub set frequent is deleted.

Step to find the minimum support in Apriori algorithm:

Min support count= number of transaction * sup count.

For example: When the percentage values of support and confidence is given 60% and 60% respectively and number of transactions are 5 then

The result will be 5*60/100 = 3.

The performance of the classical apriori algorithm decreases because the algorithm requires multiple scans and when database size is very large, then it become very time taken process.

Apriori algorithm can be implemented on various applications and discusses how effectively e-commerce application can be used with Apriori algorithm to help in business decisions by knowing the customer buying behaviour analysis especially in the retail sector. The role of Apriori algorithm is also explained for finding the itemsets which are frequent and association rules generation is also included. The dataset includes analysis of the set of products purchased by the customer in a period of time is selected. Two main measurements are used for finding the frequent itemsets and strong rules of association.

For all the transactions support is calculated which defines the association of dataset or item set [12]. This algorithm has some drawbacks. These are as follow:

- It repeatedly scans done over the database.
- The frequent item set length is directly proportional to the total database passes.
- For generating the candidate much time, resources are required.
- ARM is not so efficient for large data set.
- ARM treats all items in the database equally.

Association rule mining is employed in many attractive application areas including engineering, marketing, medicine, and more. The base paper reviews about various applications and discusses how effectively e-commerce application can be used with Apriori algorithm to help in business decisions by knowing the customer buying behaviour analysis especially in the retail sector. The role of Apriori algorithm is also explained for finding the frequent itemsets and generating the association rules.

The dataset of market basket analysis of the set of products purchased by the customer in a period of time is selected. Two main measurements are used for purpose of finding the frequent itemsets and strong association rules which are support and confidence values respectively. The support for all the transactions is calculated which defines the association of dataset or itemset [7].

Apriori includes two main steps: join step includes generating candidates by performing join operation between the frequent itemsets according to levels and prune step includes those itemsets which are discarded because support of these itemsets is less than decided support and also discard itemsets containing infrequent subsets.

Procedure for finding the frequent item-sets in Apriori algorithm:

1. All individual elements are Searched (1-element itemset) that have a minimum support of s.
2. Repeat:
    a. Search for all i+1 element itemsets that have a minimum support of itemset based on the results of previous search for i-element itemset,
    b. The set of all interesting frequent (i+1) itemsets are discovered.
3. Until item-set size reaches maximum.

Procedure for defining the mining association rules using Apriori algorithm:

- Use Apriori to generate frequent itemsets of different sizes.
- At each iteration divide each frequent itemset x into two parts antecedent (LHS) and consequent (R.H.S) which represents a rule of the form LHS→RHS
- The confidence of such a rule is support(X)/ support (LHS).
- All rules whose confidence is less than minimum confidence are discarded.

1.7Application of Data Mining

Main applications of data mining are:

1. Data Mining for Biomedical and DNA data analysis

Data Mining plays crucial role in Medical science like Biomedical, DNA, Genetics and Medicine. It is very important to understand the mapping relationship between the variation in human DNA sequences and the disease susceptibility. Data mining helps to improve the diagnosis, prevention and treatment of the diseases. Genes patterns and functions has used on large scale. So Data Mining tools helps DNA analysis and It also helps for getting different patterns.

2. Data Mining as Financial Data Analysis

Financial data is generated by banks and by other financial sectors. This data can be getting from these banks. This financial data is usually efficient, complete and has high quality. Data must be complete because insufficient data cause various problems. Data analysis can be done through one of way that should be Systematic. Data Mining is one of way doing data analysis. Data Mining is used different procedures i.e. data collection and understanding, data refinement, model building and model evaluation and deployment. These steps are very helpful to deal with analysis of financial data. Better decisions can be taken by the proper analysis of financial data.

3. Data Mining for Real Data Industry

When sale and purchased of products is increased then collection of data is also increased. Numbers of companies develop day by day. Because of development of companies, use of e- commerce is increased. When customer buy and sale products through e-commerce then experience for online shopping is also increased. Electronic commerce is a way of the buying and selling of products, services, and information through internet. E-commerce can be possible by internet.

4. Data Mining for Telecommunication Industry

Telecommunication services have increasing day by day various services. These services are local and long distance voice communication services. These services are possible by fax, pager, cellular phones and e-mails. These services are joined with various communication technologies. These technologies may be any network and others. These technologies work effectively due to advert of telecommunication. Effective results can be obtained by Integration of data mining with communication technologies.

CHAPTER 2
LITERATURE REVIEW

Frequent itemsets play crucial role in organization and online markets. Many algorithms are proposed by different authors.

[1] Jitendra Agrawal· Shikha Agrawal· Ankita Singhai· Sanjeev Sharma proposed SARIC(set particle swarm optimization for association rules using the itemset range and correlation coefficient). Apiroi is basic algorithm for finding frequent itemsets but it takes more time for generating association rules in large database. Eclat is developed to remove limitations of apriori but eclat need user defined threshold. SARIC removes limitations of apiroi and eclat. It generates association's rules with less time.

[2] Mingjun Song, and Sanguthevar Rajasekaran they proposed a novel algorithm for mining complete frequent itemsets. This algorithm is known as TM (Transaction Mapping) algorithm. In this algorithm, mapping and compression of id of each item can be done to continuous transaction intervals in a different space and compare compression coefficient with interval intersection and if it becomes smaller than the average number of comparisons for intervals intersection the algorithm switches to transaction id intersection. This algorithm is based onFP-growth and dEclat. It generated long and short items.

[3]'Sandy Moens, Emin Aksehirli and Bart Goethals' works on frequent itemset mining for big data in which distributed version of eclat algorithm is defined. Frequent itemsets from big data is found. Big data is nothing but it contains huge amount of data. So finding frequent itemset on big data is very useful. Frequent item set mining plays very important role for extracting the useful knowledge. In this paper, two algorithms are used one is Dist-Eclat algorithm which is used for speed purpose and second algorithm is BigFIM that focuses on extracting data from big databases. Hybrid approach is used with BigFIM for mining. In this paper frequent item set mining techniques are applied on the Map Reduce platform. Distributed version of eclat (Dist-eclat) divides the database into different processing units. Eclat concerns with speed based on K-FI. This k-FI is extracted through apriori algorithm and search frequent item sets using eclat. But it takes more processing time.

[4] Marghny H.Mohamed.Mohammed M.Darwieesh works on efficient frequent itemsets algorithm. Frequent itemsets are very important for various data mining tasks and for generating association rules. In this paper, there are two algorithm are used. CountTableFI algorithm and BinaryCountTableF algorithm. Both these algorithms are different from apriori and other algorithms. The main idea for developing these algorithms is to present all transactions in binary form and decimal form. Because binary and decimal form of transactions can be easily understandable. By using binary and decimal form, user can use subset and identical set properties. In Countable algorithm, original transaction data is converted into new smaller transaction data. Then it generates new 'merge transactions' where' merge transaction' is collection of various transactions and then frequent itemsets are generated by creating count of table of items. In BinaryCountTableF algorithm, original transaction dataset is represented in 0/1 form. By using 0/1 form, user can convert data into decimal form. Then all transactions are merged. User can generate frequent itemsets using transactions.

[5] Bina Kotiyal, Ankit Kumar, Bhaskar Pant, R.H. Gaudar,Shivali chahuan and Sonam Junee work on analysis of user behaviour by comparing eclat and apriori algorithm in web log. WWW provides necessary information to users on web. Web log files are created on web. These web log files stores the information about interaction between client and service provider. It also stores information about web pages which are accessed by user. The information which keeps in the web log files is used to predict the behaviour of user. In this paper, two algorithms are used to determine which pages are accesses by user. These algorithms are eclat and apiroi. Eclat is more scalable than apiroi because in eclat algorithm fewer tables are generated as compared to apriori. So less time is required to perform the analysis in eclat algorithm.

[6] S Vijayarani and P Sathya, author says that frequent pattern mining is the way of f mining useful data. It takes useful patterns and items from huge amount of data. Items should be satisfied resulted minimum support threshold. A frequent pattern is a pattern that occurs frequently in a dataset. Association rules are created after finding the frequent itemsets. Itemset which satisfied minimum support and confidence are considered as frequent itemsets. All the transaction of that database must contains frequent itemsets because similar types of items are sold daily. By using frequent itemsets, we can generate association rules, sequences rules, web log mining and helps to get interesting patterns from complex data. Ordered sequence of items is known as Data stream. It i defines

uninterrupted flow of a long sequence of data. Some real time examples of data stream data are sensor network data, telecommunication data, transactional data and scientific surveillances systems. These data makes trillions of updates every day. So it is very difficult task for storing the data. Some mining method should be used to solve this type of problems. Data mining is the way of identifying valid, original, potentially useful and ultimately understandable patterns in data. Hidden predictive information is extracted from large data base. A lots of algorithms have been used to find out the frequent item set. Apriori algorithm is the very first classical algorithm used to find the frequent item set. Apart from Apriori, lots of algorithms generated but they are similar to Apriori. They are based on prune and candidate generation. It takes more memory and time to find out the frequent item set. In this paper, they have studied about how the clat algorithm is used in data streams to find out the frequent item sets. Éclat algorithm need not required candidate generation.

[7] Shaobo Shi.Yui Qi,Qin Wang work FPGA acceleration for intersection computation in frequent itemset mining in which performance of eclat algorithm is increased. Huge amount of "sorted-set intersection computation" decrease the performance of algorithm. FPGA is one of the platforms which can be applied on parallel data.

[8] Guo-Cheng Lan, Tzung-Pei Hong, Hong Yu Lee introduced weighted frequent itemsets concepts. In Weighted frequent itemsets, weight values are set for all items. Weighted frequent itemsets mining improve data mining techniques. Transactions upper bound model is used to find upper bound. Upper bound is highest weighted value of items.

[9] PENG JIan , WANG Xiao-ling proposed an improved association rule algorithm according to itemset matrix and cluster matrix. Itemset matrix decreases time of comparison that is made between candidate itemsets and records. And Cluster itemsets makes cluster of records. When database is changed then only changed part is scanned. There is no need to scan full database again.

[10] D. Gunaseelan, P. Uma, author proposes many approaches those are used to overcome the drawbacks of the Apriori algorithm as to improve its efficiency. The proposed approach is presented which decreases pruning operation of candidate item-set. Data consistency is improved by resolving the problems like bad data or duplicate data and instead of finding the whole dataset, focusing on finding association in the filtered

dataset. Other technique is also used to overcome the exceed limit of memory size by both frequent and infrequent item-sets by dividing dataset into horizontal partitions. The proposed idea reduces the size of each transaction and takes less time in comparison to Apriori algorithm to handle the data.

[11] Jaishree Singh, H. R, author proposes an approach for candidate item-sets generation which overcome wasting time for scanning the whole database as in Apriori algorithm. Before scanning all transaction records, use only before generated frequent item-set(Li) to get the transaction IDs of the minimum support count between X and Y items. Repeat the steps to identify all frequent item-sets. The candidate support count generation in improved algorithm is less time consumed as compared with the Apriori algorithm.

[12] Chanchal Yadav, Shuliang Wang, Manjot Kumar, they describe an Improved Apriori algorithm which removes the unnecessary item sets and thus reducing the time for scanning the database. This proposed algorithm also reduces the generation of similar sub-items in the candidate item sets pruning step. The apriori algorithm performance is also improved in terms of speed for mining association rules. The improved algorithm has some limitations also as this is not efficient algorithm to manage the new database.

[13] kan sin works on new algorithm for discovering associations rules that is enlightened by éclat algorithm called LOGeclat. Logeclat find frequent pattern by using special candidate. Special candidate continue update data and reduce time. Logeclat is combination of éclat and special candidate method. It extracts correct itemsets. It takes less time then éclat but numbers of iterations are more.

[14] M.N. Noothir proposed Log server files are very useful for each organization network. It develops large amount of data daily. The data contains useful and meaningful information. This information is very difficult to understand. So web log analysis tool is used for understanding information. In this paper Apriori algorithm is considered for frequent itemsets generation and UMPNA tool is used. UMPNA tool cleans the data. By using this tool, network performance is increased.

[15] Sang Lin; Hu-yan Cui; Ren Ying; Zhou-lin Lin say that Association rule mining helps for finding all association rules between all the item sets in the database. Support and confidence must satisfy in the algorithm used in the paper. There are numbers of

constraints like knowledge constraints, data constraints but main constraint is item constraints. “B constraint” is major constraint. “B Constraint” is a predicate for power set of collection of itemsets. “B Constraint” can be divided into correspond classes. And before applying constraints on eclat algorithm, sorting of all items is done. When all constraints are applied then re-sort all items is done. It is mainly used for dealing with others constraints. In this paper, only maximal item set are found which satisfied all the constraints. It is very helpful in finding constraint frequent item sets in long pattern database.

[16] Mahanti , Aniket and Reda Alhajj proposed visual interface for displaying the result of eclat and apriori algorithm. Visual interface is very user friendly interface. User can easily understand it. By this interface, user can understand the frequent item sets. User can judge that which item is mostly used. There is various application of association rule mining. Main application is market basket analysis. Market basket analysis is mostly required in market, malls, and online markets. There are other applications available such as customer profiling, fraud detection, credit risk analysis and so on.

CHAPTER 3
PRESENT WORK

The association rule generation is the main challenge in data mining for finding frequent data items. The various algorithms had been proposed in data mining. In this work, we are focusing on eclat algorithm for generating associations' rules in data mining. In previous times various enhancements had been proposed in eclat algorithm for association rules generation. In the previous work, the huge dataset will be processed in parallel manner and most frequent data items will be discovered in data mining. The minimum support count and confidence level also calculated with the enhanced formula. These proposed techniques will increase efficiency of eclat algorithm. In this work, we will reduce number of transaction for association rule generation in eclat algorithm. The enhancement will be based on to convert dataset into transposed dataset and apply top down scan to search elements from the dataset. This enhanced technique will reduce number of transactions and reduce escape time of the algorithm for association rule generation in data mining. Market basket analysis is helpful in insurance and medical. In insurance companies, link analysis is done for fraud. In medical, symptoms are analyzed. Frequent item sets helps to increase supply of goods in organization according to customer demands. It helps for taking decisions in the organization. It improves business strategy. Frequent item sets are used in business market and shopping malls. Large amount of data is stored in the database. When organizations want to get data they can access data according to need. They do not need to fetch all the data. They mostly find the frequent data from database. There is another way on which we can focus is 'data structure'. Data structure helps to contain item sets which are intermediate candidate. By this stored data can be easily accessed. Frequent item sets algorithm can be used in parallel mining. We can reduce the time complexity if parallelism is applied when calculating frequent item sets

CHAPTER 4
OBJECTIVES OF STUDY

Objectives are the most vital part of our research which helps us to achieve our desired aims and expectations from the research. It is said that if objectives are clearly defined that half part of research is considered to be done. In this research, we are going to remove limitations of eclat algorithm.

The Objectives of this research work is :

1. Escape time of proposed algorithm will be reduced.
2. Complexity of Proposed Algorithm will be reduced.
3.Eclat Algorithm requires more number of iterations for generating frequent itemsets.This limitation is removed in proposed algorithm by transposing of original database.
4. The proposed algorithm provide easy way to find the items in market basket dataset.
5. The proposed enhancement will be implemented in MATLAB by taking market basket dataset .

CHAPTER 5

RESEARCH AND METHODOLOGY

5.1 Eclat algorithm

Eclat algorithm finds the items from bottom like depth first search. Eclat algorithm is very simple algorithm to find the frequent item sets. This algorithm uses vertical database. It cannot use horizontal database. Horizontal database is converted into vertical database.

ALGORITHM:

Input: $F_k = \{I_1, I_2,,,, I_n\}$ // cluster of frequent k-itemsets.

Output: Frequent l-itemsets.

Bottom-Up (F_k) {

1. for all I_i € F_k
2. F_k +1 = ∅;
3. for all I_j € F_k , i < j
4. N = I_i ∩ I_j;
5. if N .supı>= min_sup then
6. I_j € F_k = F_k +1 U N;
7. end
8. end
9. end
10. if F_k +1 ≠ ∅ then
11. Bottom-Up (F_k +1);
12. end

13.}

As given in Algorithm, take F_k as input. Output will be Set of frequent itemsets. In Bottom-Up (F_k),for loop is defined in which all items belong to database F_k under first step. In step2, take F_k +1 as empty database. In step3,check that item exists in the database F_k or not. If item exists in the database F_k then calculate support of items in step4 where support means how many times item occurs in database F_k. In step5, compare support of items individually with minimum decided support. In step6, put those items whose support is more than then minimum support in database F_k +1. In step7, check that database F_k +1 is empty or not. If it is not empty then start the same procedure for other database.

Example of Eclat algorithm

Table 1: Original Database

Transaction	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5
101	0	0	1	0	0
102	0	1	1	1	0
103	1	1	0	1	1
104	1	1	0	1	1
105	1	0	0	1	0
106	0	1	1	1	1
107	1	0	1	0	0
108	0	0	1	1	1

In table1, 8 transactions are taken. 5 items are sold in 101 to 108 transactions. 0 represent item does not exist in the database and 1 represent item exists in the database.Table2 shows support of each item in every transaction i.e items comes in how many transactions

and compare support of each item with decided support. Table3 shows only those items whose support is equal to and greater than decided support. In table4, we have paired items of table3 and calculate their support i.e in total numbers of transactions in which their pairs comes collectively. Table5 shows only those paired items whose support is equal to and greater than decided support. In table6, we have paired items of table5 and calculate their support i.e in total numbers of transactions in which their pairs comes collectively. Table7 shows only those paired items whose support is equal to and greater than decided support.

Table 2: **Calculating Support**

Itemset	**Support**
ITEM1	4
ITEM2	4
ITEM3	5
ITEM4	6
ITEM5	4

Average support is 3. Take items which has support equal to or more than 3.

Table3: **Table after comparing Support**

Itemset	**Support**
ITEM1	4
ITEM2	4
ITEM3	5
ITEM4	6
ITEM5	4

Table4: Paired Database

Itemset	**Support**
ITEM1,ITEM2	2
ITEM1,ITEM3	1
ITEM1,ITEM4	1
ITEM1,ITEM5	2
ITEM2,ITEM3	2
ITEM2,ITEM4	4
ITEM2,ITEM5	3
ITEM3,ITEM4	3
ITEM3,ITEM5	2
ITEM4,ITEM5	4

Average support is 3.

Table5: **Paired Database after checking support**

Itemset	**Support**
ITEM2,ITEM4	4
ITEM2,ITEM5	3
ITEM3,ITEM4	3
ITEM4,ITEM5	4

Table6: Paired Database

Itemset	Support
ITEM2,ITEM4,ITEM5	3
ITEM2,ITEM3,ITEM4	2
ITEM3,ITEM4,ITEM5	2

Table7: Result Database

Itemset	Support
ITEM2,ITEM4,ITEM5	3

Eclat algorithm finds the frequent item sets. Eclat is based upon depth first search. Eclat is used bottom up approach. When vertical layout is used then more numbers of iterations are generated. So we try to improve the problem. Equivalence class clustering and bottom up lattice traversal is known as ECLAT algorithm. It uses vertical data layout. It use tid set intersection that is transaction id intersection to compute the support of a candidate itemset for avoiding generation of subsets that does not exist in the prefer tree.

1. To calculate 1st candidate frequent item sets, it assume vertical t-id list database. Each time it read its corresponding tid list from given database and increment items support for each entry. It improves fastness of the support computation.
2. It scans database only once and creates the vertical database which identify each item in the list of transaction
3. With the help of the eclat algorithm. We can put tid to each transaction. This algorithm checks the transaction data item wise. It puts id to each transaction.
4. Now set up the items with their support count. To find out frequent items, support count of each item is necessary. With the help of the support count we can calculate frequent item set.
5. The support count of an item set is greater than equal to min. support threshold that item is said to be frequent items.

In the figure 5.1.1, Whole process of eclat algorithm is defined. In this figure, we first take dataset as input. After that pre-processing of dataset is done. By Pre-processing, all the unwanted data, empty sets and errors are removed from the dataset. Then we apply eclat algorithm for frequent itemsets generations. First we convert original datasets into vertical dataset. We use bottom up approach for searching the items in dataset. At last we get frequent itemsets. Figure 5.1.1 shows the flow chart of eclat algorithm. It tells that how eclat algorithm works and how associations rules are created with the help of eclat algorithm.

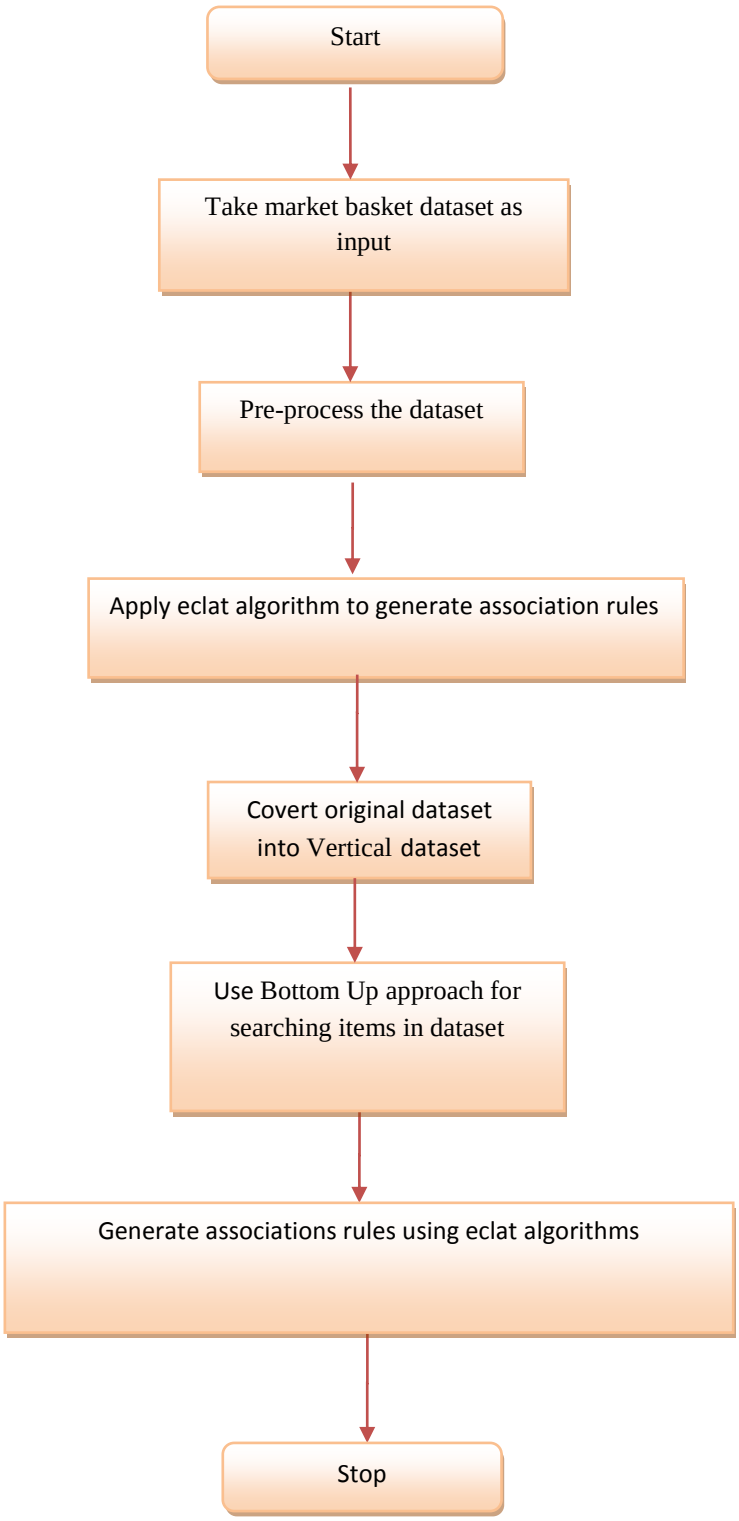

Figure 5.1.1 Flowchart of Eclat algorithm

5.2 PROPOSED AGORITHM

PROPOSED AGORITHM:

Input: Data set is in Transposed form contain transactions and items

Output: FI as Frequent Itemset

1. Initialize: P =(T(i),T(j)) for all items in dataset T [Here T(i) is number of rows and T(j) is the number of columns in the dataset T]
2. [row column]=size(P) to get the number of rows and columns in the dataset P
3. While(P(I, j)=[rows columns])
4. for(i=0;i=P(rows); i++) [it is the for loop to count number of items]
5. for (j=0;j=P(columns);i++)
6. If (p(i,j) =1)
7. Count=count+1;
8. Else
9. Count=count;
10. End
11. End
12. End
13. if (count>=support)
14. P(i,j)=P((i-1),P(j-1))U(P(i),P(j);
15. Else
16. P(i,j)=P(i,j);
17. End
18. FI=P((i-1),P(j-1))U(P(i),P(j));
19. end

In this algorithm, numbers of items are taken as input. Output will be set of frequent itemsets. First Initialize P as one dataset which contains the information about rows and columns of transposed dataset. Then put all information of dataset P into [row column].

Then checks support of all items and compare it with minimum support. The items which satisfied minimum support are taken as frequent itemsets .

Eclat algorithm is used to find the associations rules. In previous, Eclat algorithm uses vertical form for storing items and it uses bottom up approach. But in previous techniques, accuracy is major issue. So we will introduce new algorithm which will increase accuracy, so by using enhanced Eclat algorithm, we can find frequent items which are helpful for various companies, organizations and online markets. So this enhanced Eclat algorithm provides better result with less time. Time to process these items will be reduced by enhanced algorithm. There will be less number of iterations required for process these items and escape time is also reduced. This algorithm takes less time. Transposition of original database reduces escape time. It reduces number of iterations. Thus complexity of proposed algorithm is reduced.

Example of Proposed Algorithm

In table8, take transpose of table1 (Original Database). Calculate support of each transaction in table9 and then compare support of each transaction with minimum support. Take those transactions in table10 which satisfy minimum support. Make pair of transactions and calculate support of each pair in table11. Take all pairs in Table12 which satisfy decided support. Take three transactions as a pair in table13 and calculate support of each pair. Pairs which satisfy minimum support are taken in table14 (Final result database).

Table8: Transposed Database

ITEM	**TID**	**TID**	**TID**	**TID**	**TID**	TID
ITEM1	103	104	105	107		
ITEM2	102	103	104	106		
ITEM3	101	102	106	107	108	
ITEM4	102	103	104	105	106	108
ITEM5	103	104	106	108		

Table9: Support Database

TID	Support
101	1
102	3
103	4
104	4
105	2
106	4
107	2
108	3

Average support is 3. So we take those Transactions whose support is 3 or more than 3.

Table10: Support Database after support checking

TID	Support
102	3
103	4
104	4
106	4
108	3

Table11: Paired Database

TID	Support
102,103	2
102,104	2
102,106	3
102,108	2
103,104	3
103,106	3
103,108	2
104,106	3
104,108	2
106,108	3

Average support is 3. So we take those Transactions whose support is 3 or more than 3.

Table12: Paired Database after checking Support

TID	Support
102,106	3
103,104	3
103,106	3
104,106	3
106,108	3

Table13: Paired Database

TID	Support
102,103,106	2
102,104,106	2
102,106,108	2
103,104,106	3
103,106,108	2
104,106,108	2

Minimum support is 3. Take those transactions which are above or equal to minimum support.

Table14: Result Database

TID	Support
103,104,106	3

In the figure 5.1.2, Whole process of enhanced eclat algorithm is defined. In this figure, we first take dataset as input. After that pre-processing of dataset is done. By Pre-processing, all the unwanted data, empty sets and errors are removed from the dataset. Then we apply enhanced eclat algorithm for frequent itemsets generations. First we convert original datasets into Transposed dataset. We use Top Down approach for searching the items in dataset. At last we get frequent itemsets.

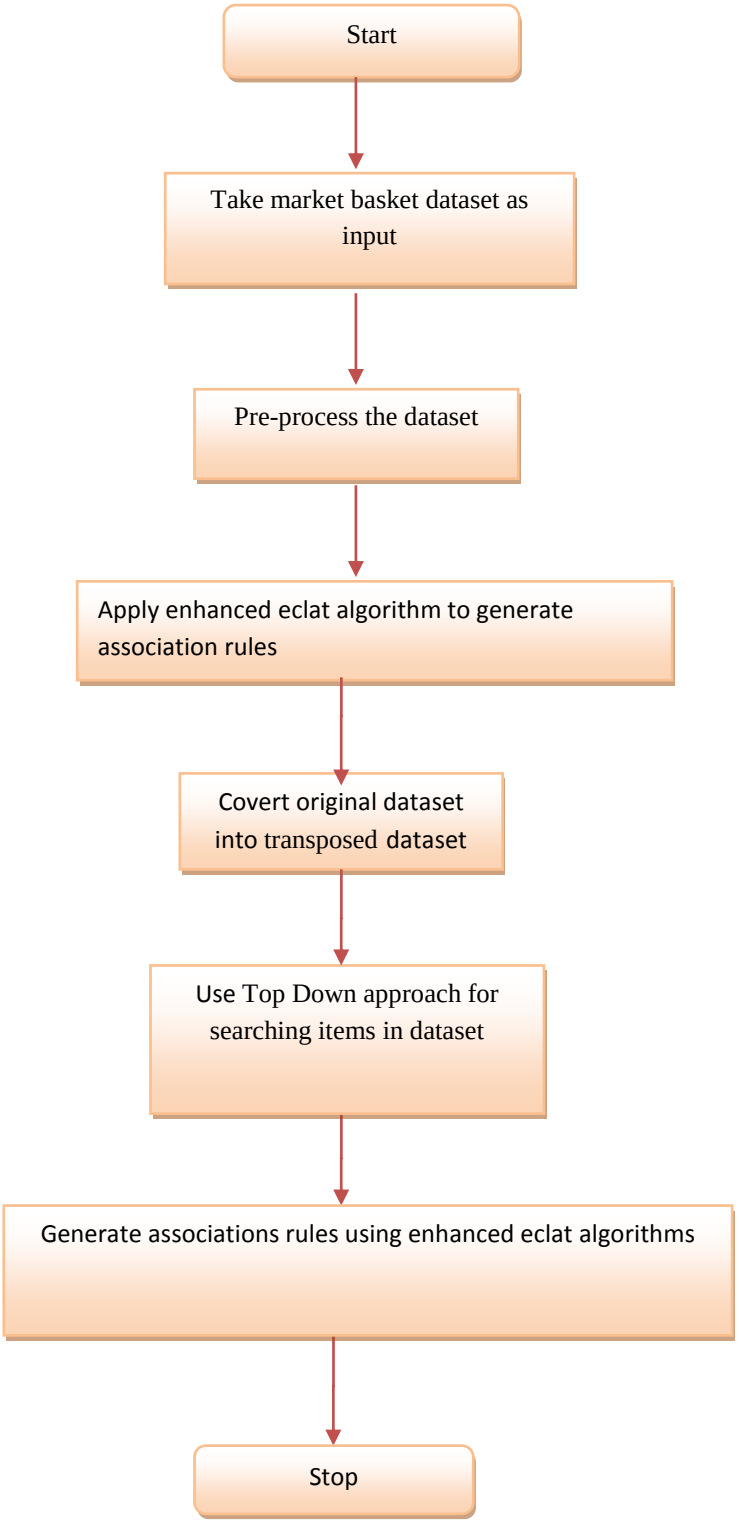

Figure 5.1.2 Flowchart of Enhanced Eclat algorithm

In this algorithm, we will use the top down technique of searching like breath for search We reduce time complexity by using top down technique. When we use bottom up approach then complexity is increased. We will use top down technique like breadth first search. When we use vertical dataset, we need to access more iteration for finding frequent itemsets we will use transposed dataset so we need fewer iteration for generating frequent item sets. . In the traditional eclat algorithm they use the vertical database and bottom up technique for element searching.

We will take market basket analysis dataset as input. Next step is to preprocess the dataset. We will apply eclat algorithm. That we will generate association rules which mean that we find those itemset which are frequently occurring in database and we are using horizontal form of data for storing the data. We will apply traditional data algorithm and enhanced algorithm then will compare the result of both algorithm in the term of performance. Performance can be measured in term of escape time and number of iterations. We will show result of both algorithms. We will prove that enhanced algorithm reduce the number of iterations. There may large amount of database. In that database there are numbers of items. We will process all the items. Time to process these items will be reduced by enhanced algorithm. There will be less number of iterations required for process these items and escape time is also reduced. After that we will get frequent items. Output is set of frequent itemsets. We will use MATLAB for implementing enhanced algorithm. Comparison of the both the algorithm is shown in the following flow chart.

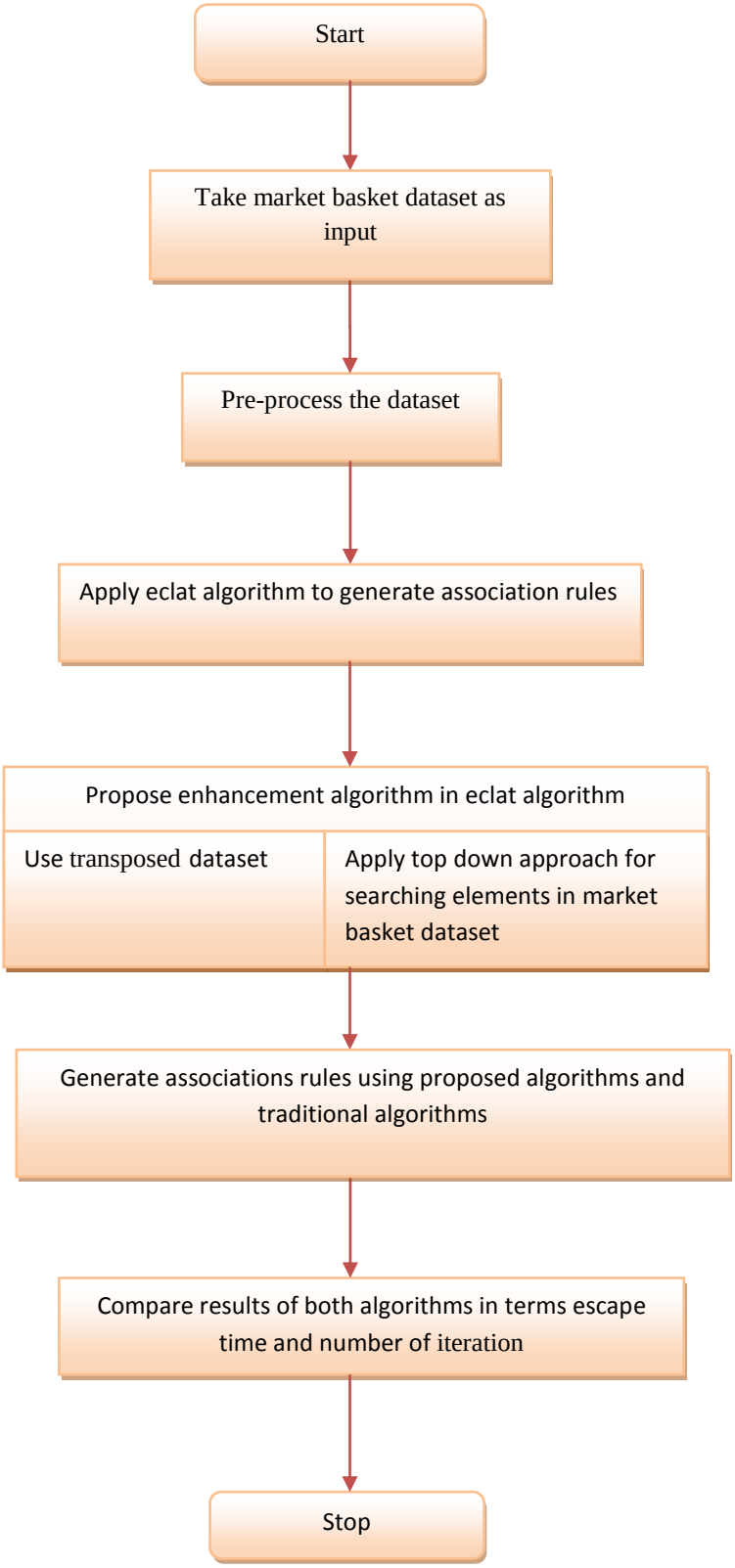

Figure 5.1.3 Flowchart of Research Methodology

CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Introduction to MATLAB

MATLAB is a high-level language and interactive environment by which we can perform computationally intensive tasks with more speed traditional programming languages such as C, C++, and FORTRAN [Mat lab Toolbox]. Or **MATLAB** (matrix laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language.

Mat lab has following functions:

1. Introduction and Key Features

2. Developing Algorithms and Applications

3. Analyzing and Accessing Data

4. Visualizing Data

5. Performing Numeric Computation

6. Publishing Results and Deploying Applications

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, numeric computation. It helps to solve more complex problems that are very difficult with traditional programming languages, such as C, C++, and FORTRAN. Add-on toolboxes (collections of special-purpose MATLAB functions, available separately) extend the MATLAB environment to solve particular classes of problems in these application areas. MATLAB gives various feature to simplify the task. MATLAB code can integrate with other languages and applications, and distribute MATLAB algorithms and applications.

Some features of MATLAB

1. MATLAB matrices and vectors

2. Dense matrices and vectors

3. Range operator

4. Size and shape operator

5. MATLAB arithmetic operators

6. Array operations

7. Back slash operator

8. Complex arithmetic

9. MATLAB software

10. Linear algebra

11. Non linear functions

12. Ordinary differential equations

13. Fourier Transformations

MATLAB Graphics

1. Plotting data(x, y)

2. Plotting data(x, y, z)

3. Movies

4. Saving Post script graph

MATLAB data handling

1. Importing ASCII data

2. Exporting ASCII data

Programming hints

1. Some basics

2. M files

3. Good Practices

Examples of problems solved with MATLAB

1. Partial differential equations

2. Matrix examples

Some more advanced examples

1. Using c++ with MATLAB

Advantages of MATLAB

1. It performs numerical calculations and visualized result programming

2. It provides graphics easily and produces code efficiently

Disadvantages of MATLAB

1. Because MATLAB is an interpreted language, it can be solved or not solved

2. Poor programming practices

6.2 Implementation of Proposed Work

Figure 6.1.1 shows the main interface in which results are shown. This interface further divides into two parts. First is time calculation part and second is complexity calculation part. In the first part, calculate the time of old eclat algorithm and advanced eclat algorithm. In second part, the complexity of old eclat algorithm and advanced eclat algorithm is calculated. There are different levels of association rules from which any level of association rule can be selected and according that selected level of association rule, time and complexity of old and advanced eclat algorithm will be calculated. In this interface, five levels of association rules are considered. First level contains C1 where C1 is collections of transactions with their support values. Second level contains L1 where

L1 is collections of transactions which satisfy decided support. First level contains C2 where C2 is collections of pairs of transactions with their support values. Fourth level contains L2 where L2 is collections of pairs of transactions which satisfy decided support. Fifth level contains C5 where C5 stores final results.

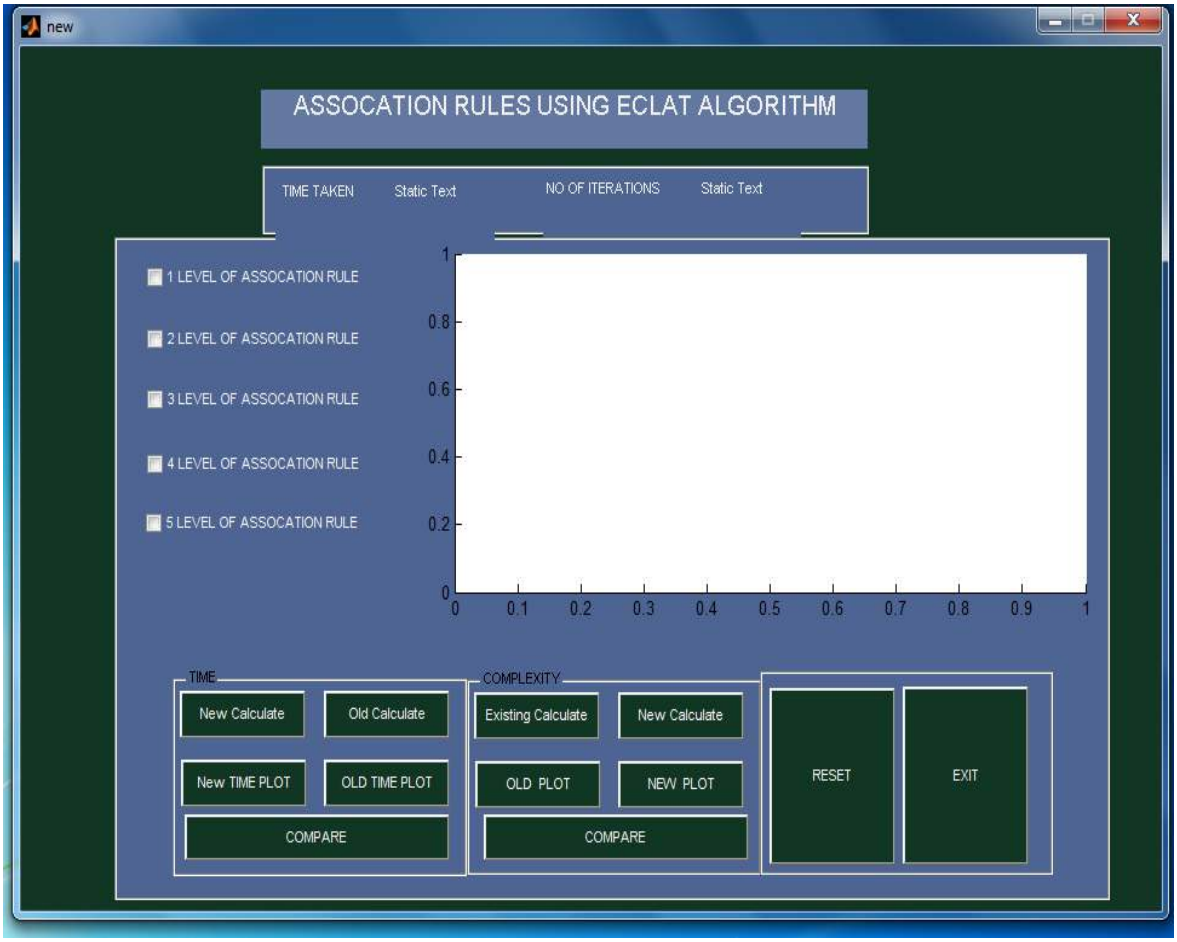

Fig 6.1.1: Main interface

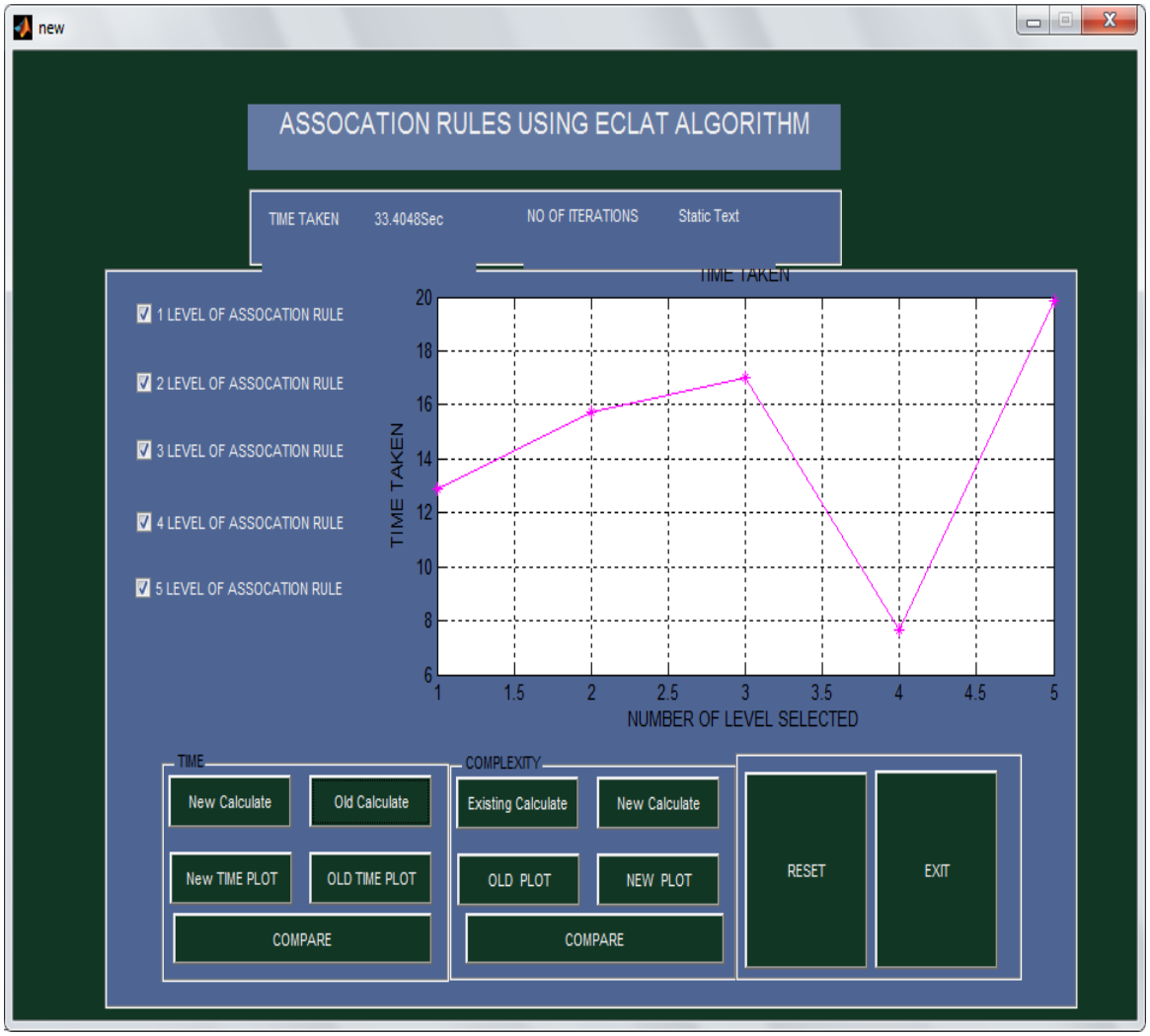

Fig 6.1.2: Time analysis of existing algorithm

As illustrated in figure 6.1.2, the various level of association rules have been shown. In this figure, time had been calculated to generate association rules using basic eclat algorithm with vertical scan and graph is shown correspond to level of association rules. By click on old calculate button under time part, we can calculate time of eclat algorithm

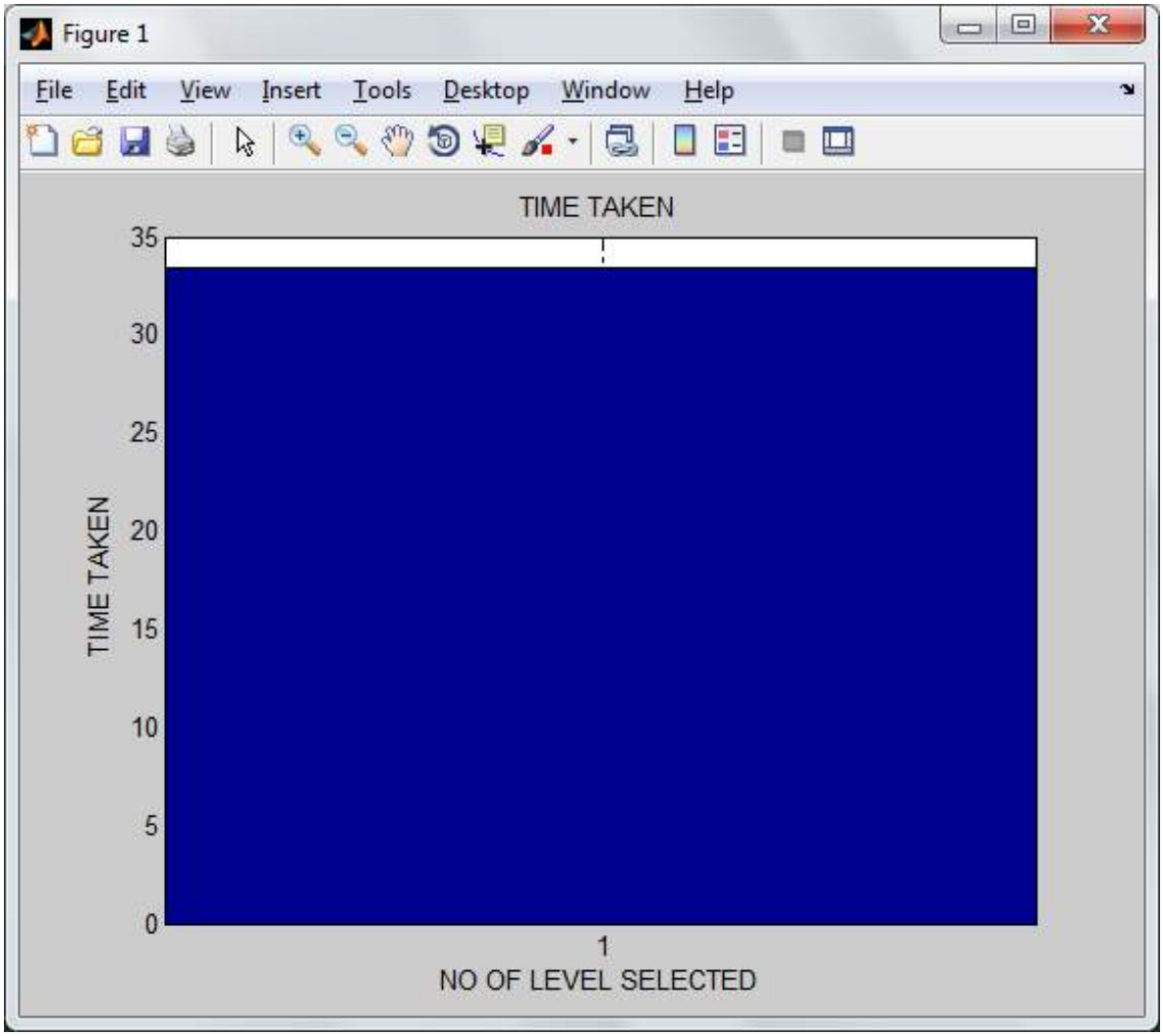

Figure 6.1.3 Individual Graph of time taking by eclat algorithm

As illustrated in figure 6.1.3, the eclat algorithm will be implement using bottom-up parsing and using vertical scan of database. In this bar graph is shown that how much algorithm will take to create association rules. This bar graph is created by clicking on old time plot button.

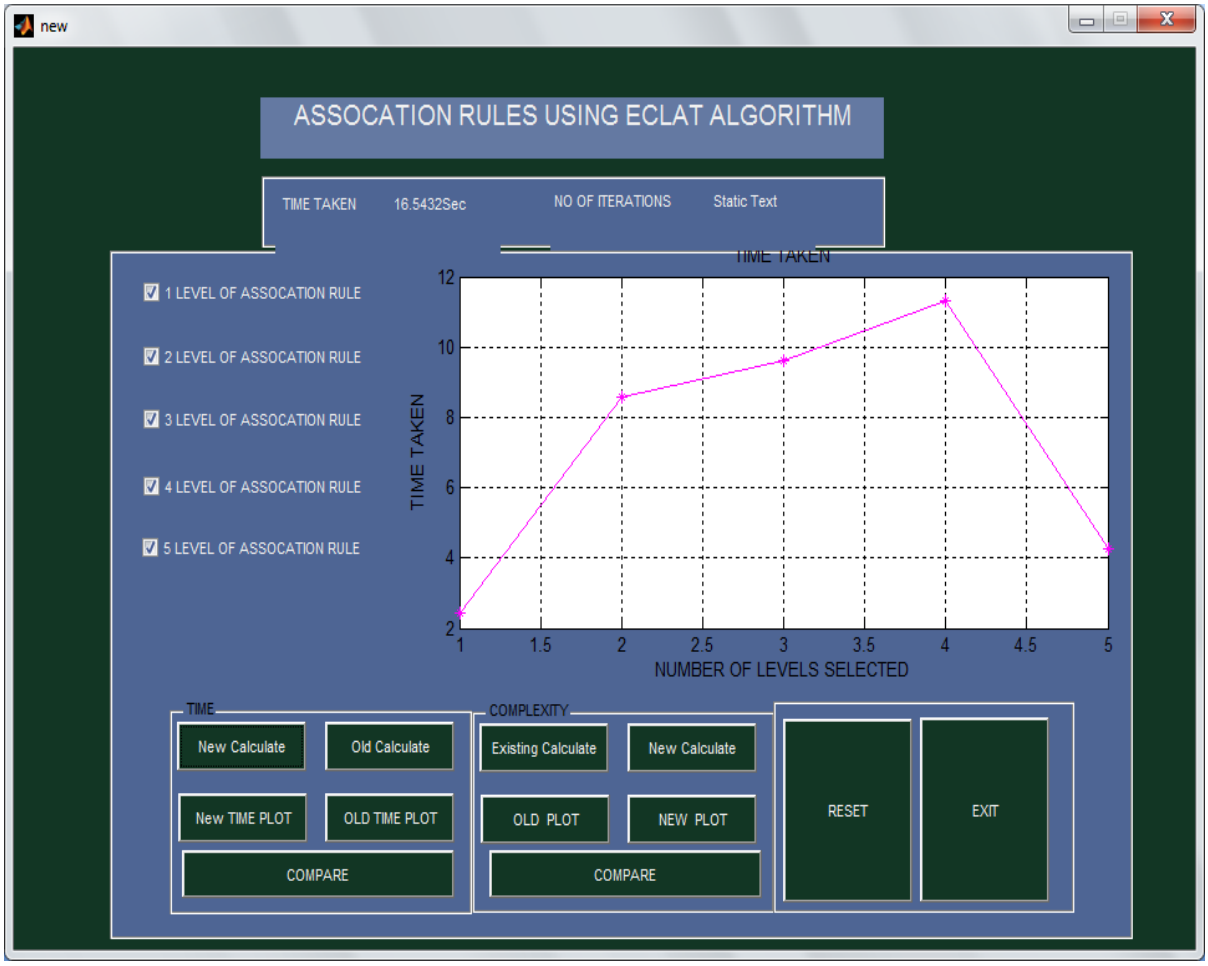

Fig 6.1.4: Time taken by enhanced algorithm

As shown in figure 6.1.4, the enhancement will be proposed in eclat algorithm to reduce processing time of the algorithm. In enhanced algorithm, the horizontal scan will be proposed and top-down approach is implemented for parsing. This will reduce the time to create processing to algorithm to create final association rules. By clicking on new calculate button, we can check time of enhanced algorithm.

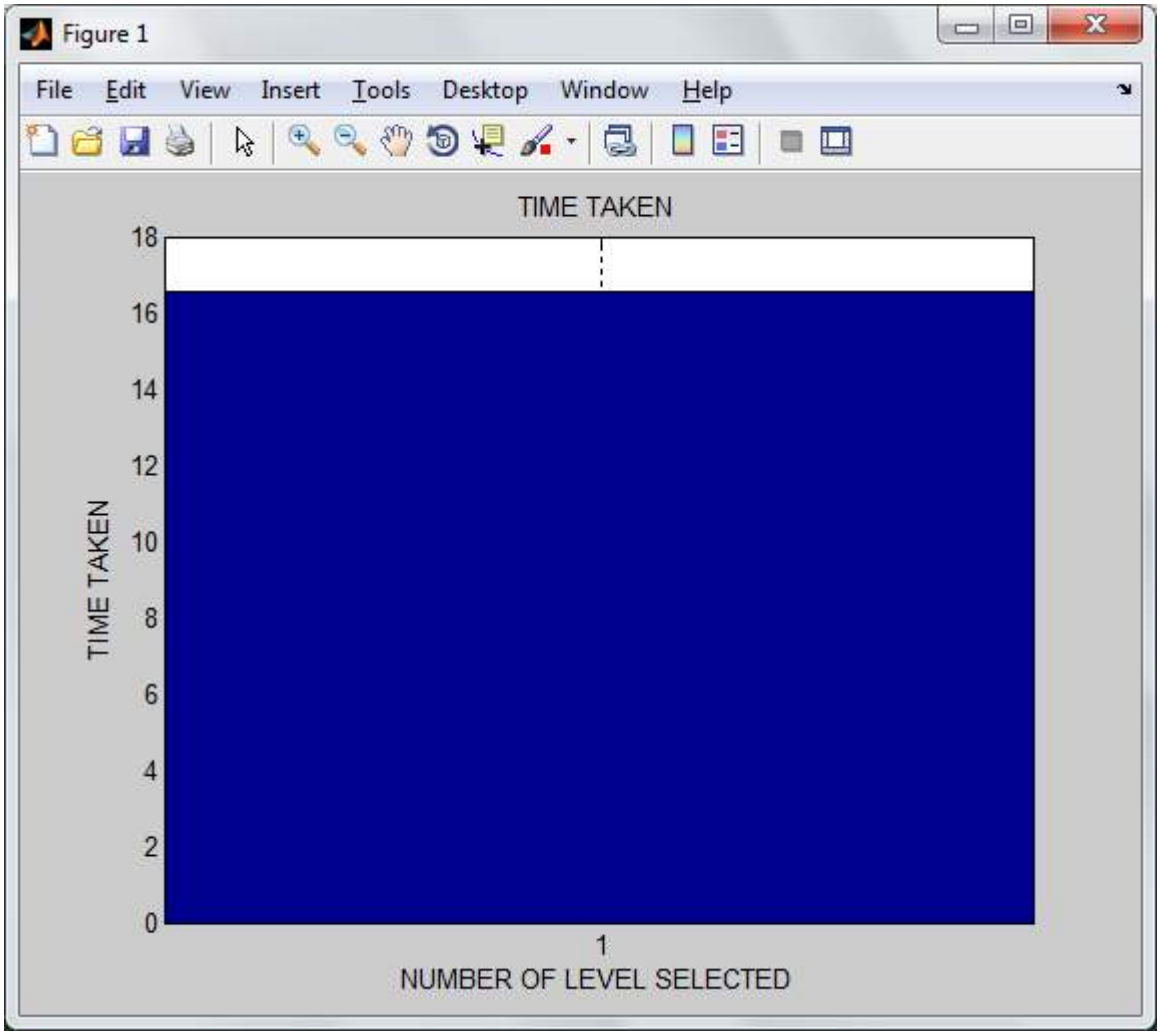

Fig 6.1.5: Enhanced algorithm Individual time graph

As shown in figure 6.1.5, the enhanced algorithm is proposed in which the transpose of the original dataset is taken and for data parsing top-down technique is implement which reduce processing time of the algorithm as shown in the bar graph.

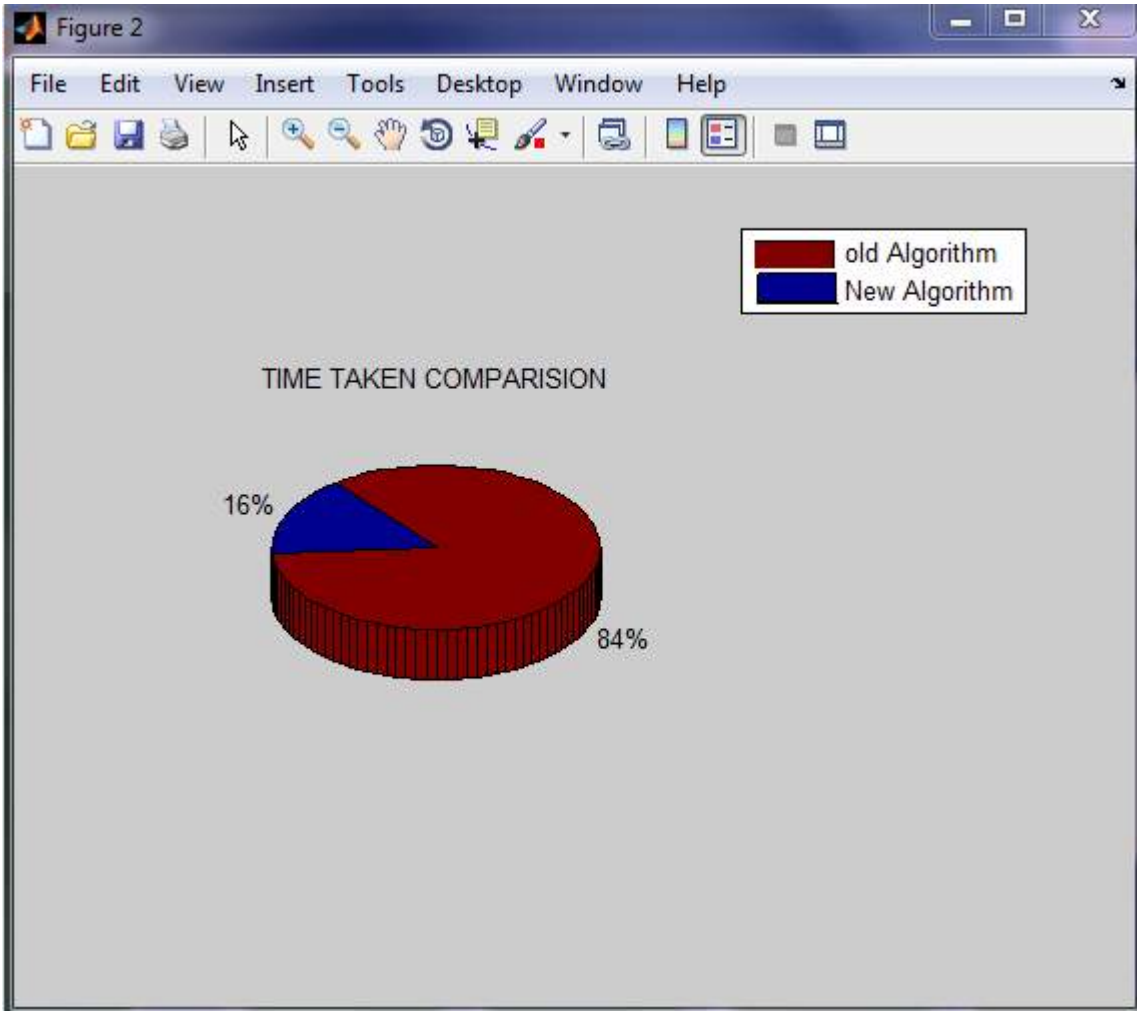

Fig 6.1.6: Time comparison

As shown in figure 6.1.6, the two algorithms had been implemented, the first algorithm is basic eclat algorithm in which vertical database and bottom-up technique is used for database scan. In second algorithm which is the enhancement of basic eclat algorithm in which transposed database is taken and top-down technique is implement which reduce the processing time as shown in the pie chart. The red portion shows the processing time of basic algorithm and blue portion shown the processing time of enhanced algorithm.

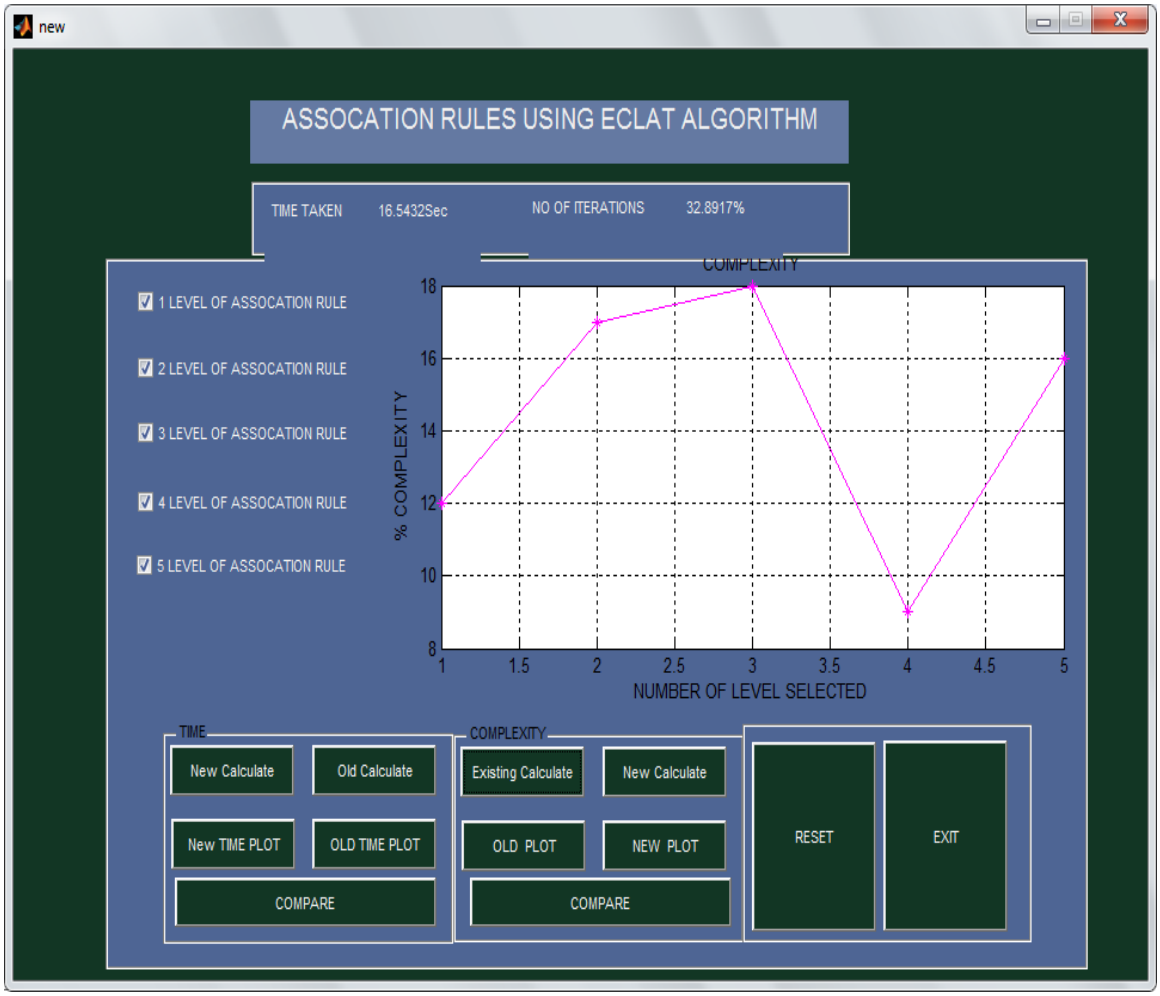

Fig 6.1.7: Basic eclat algorithm in terms of complexity

As illustrated in figure 6.1.7, the various level of association rules have been shown. In this figure complexity had been calculated to generate association rules using basic eclat algorithm with vertical scan and graph is shown correspond to level of association rules

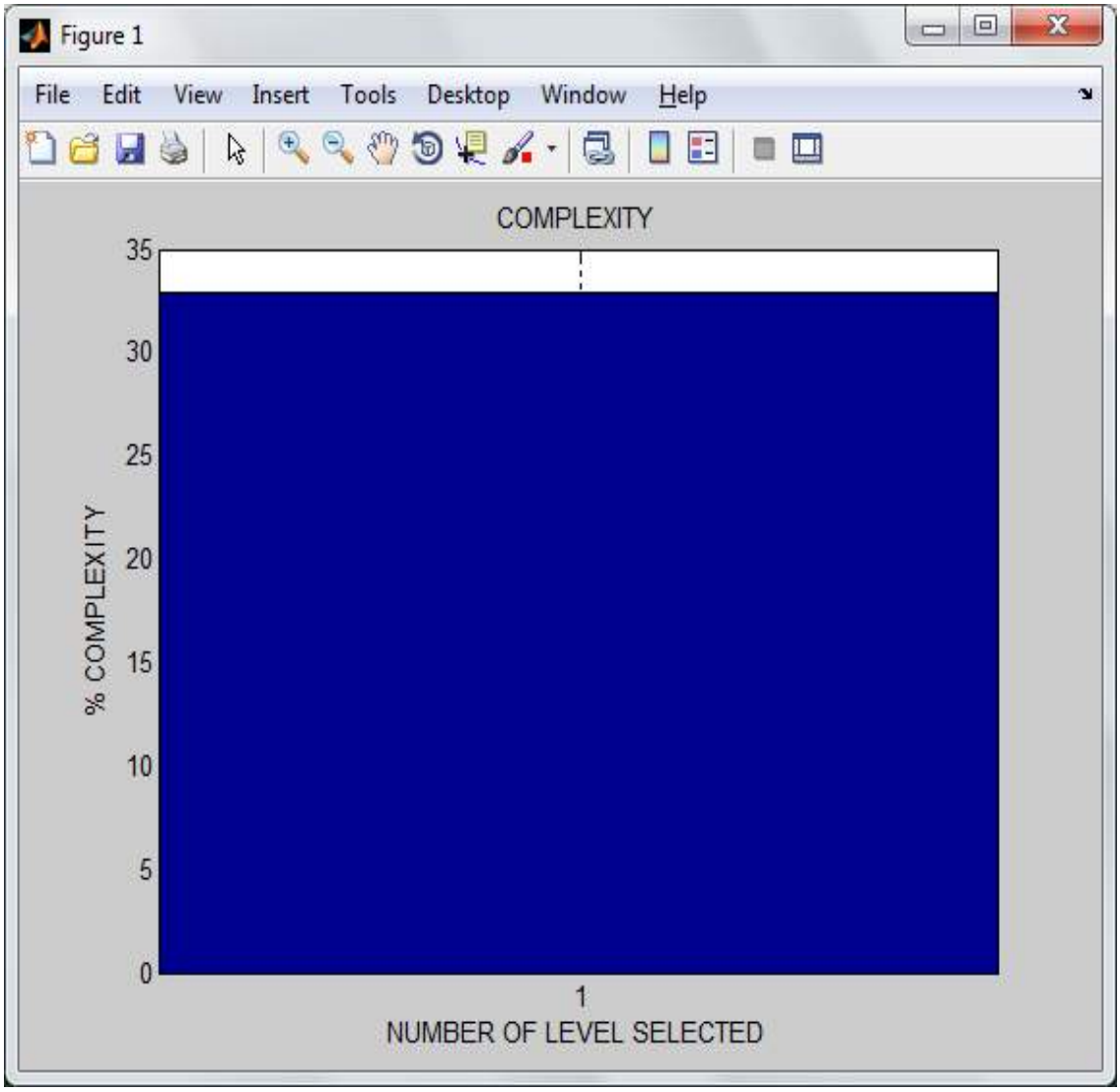

Fig 6.1.8: Individual Graph of Complexity of old algorithm

As illustrated in figure 6.1.8, the eclat algorithm will be implement using bottom-up parsing and using vertical scan of database. In this bar graph is shown that complexity of algorithm to create association rules.

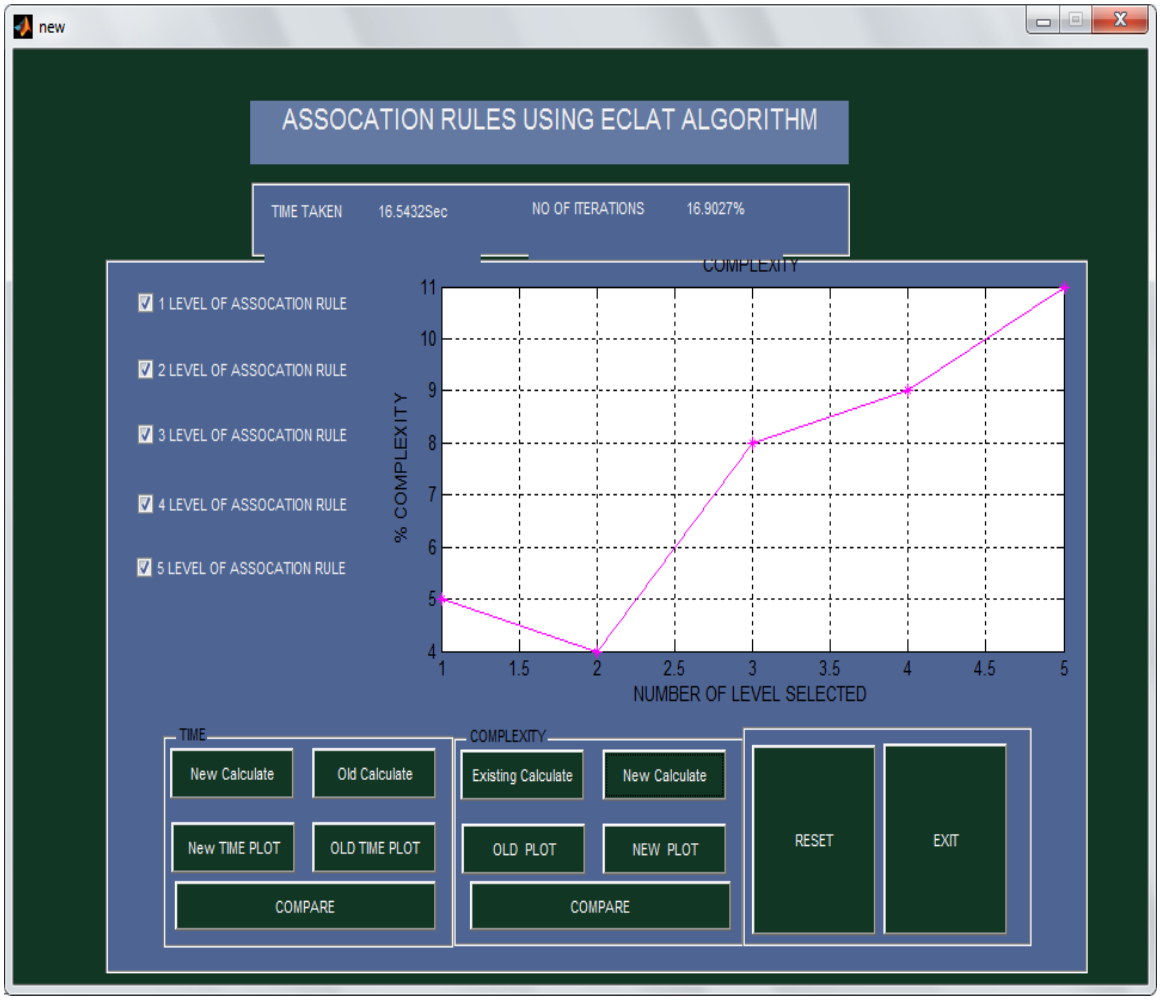

Fig 6.1.9: Complexity of enhanced algorithm

As shown in figure 6.1.9, the enhancement will be proposed in eclat algorithm to reduce complexity of the algorithm. In enhanced algorithm, the horizontal scan will be proposed and top-down approach is implemented for parsing. This will reduce the complexity to create final association rules.

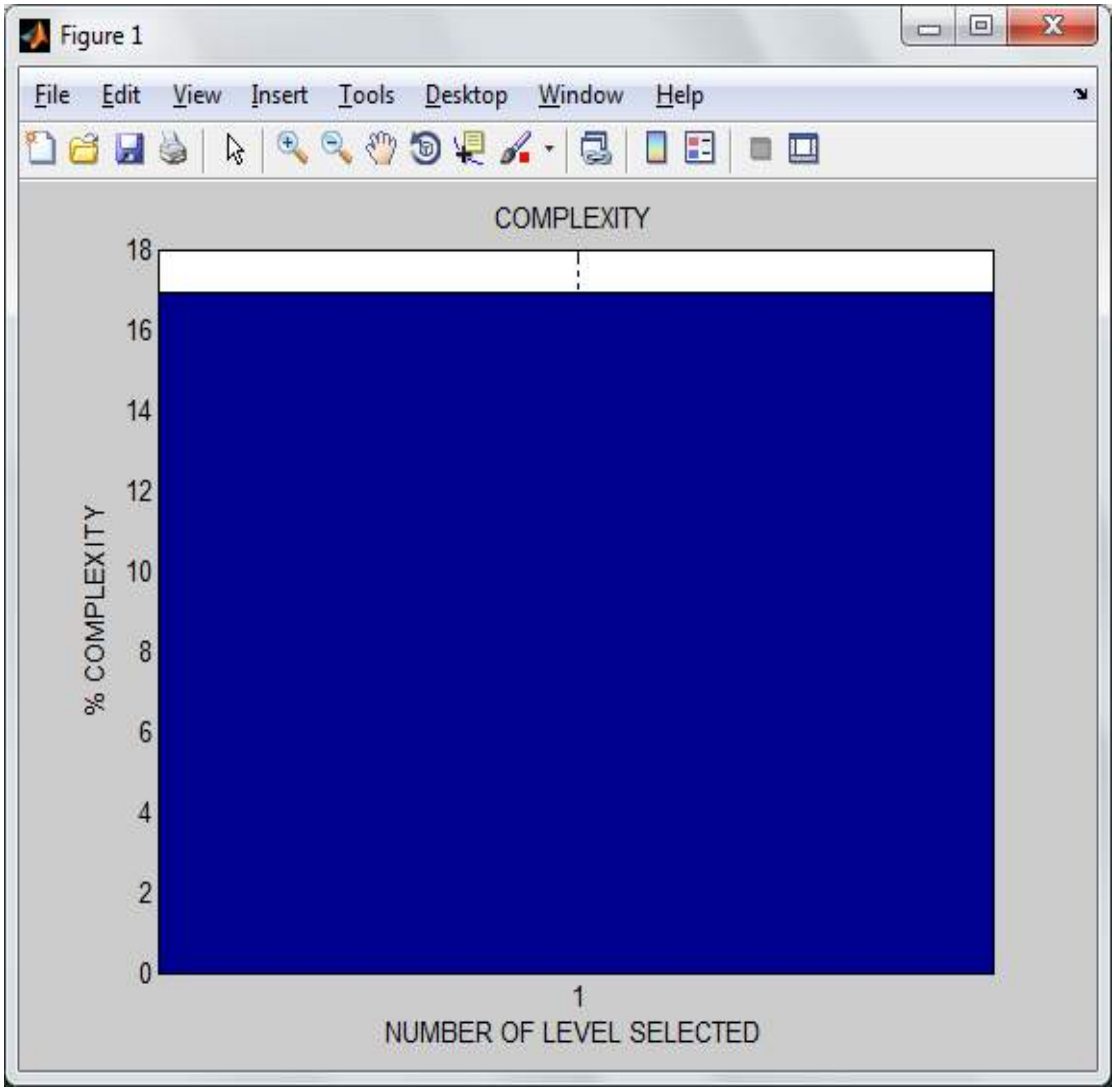

Fig 6.1.10: Individual Graph of Complexity of proposed algorithm

As shown in figure 6.1.10 the enhanced algorithm is proposed in which the transpose of the original dataset is taken and for data parsing top-down technique is implement which complexity of the algorithm as shown in the bar graph

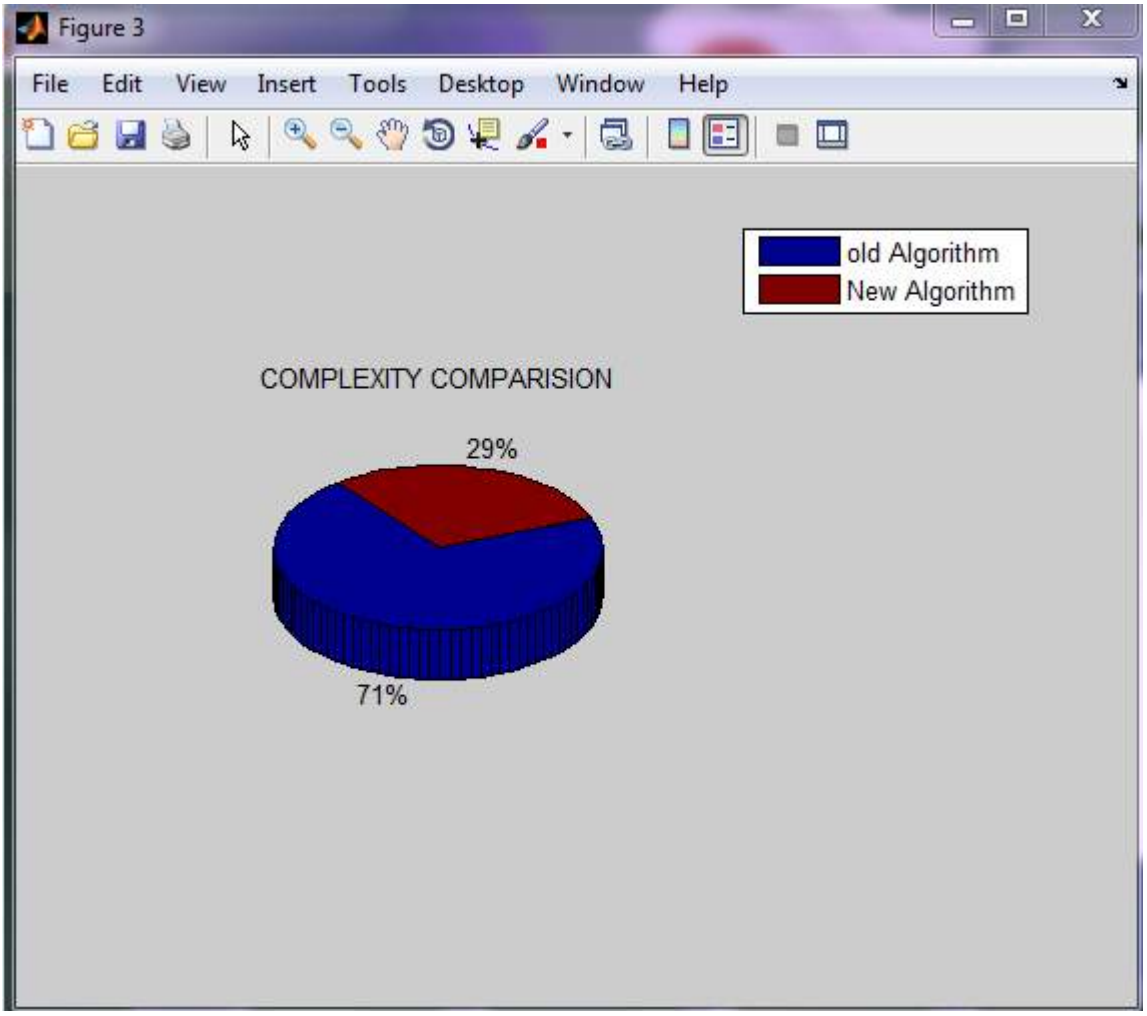

Fig 6.1.11: Complexity comparison

As shown in figure 6.1.11, the two algorithms had been implemented, the first algorithm is basic eclat algorithm in which vertical database and bottom-up technique is used for database scan. In second algorithm which is the enhancement of basic eclat algorithm in which horizontal/transposed database is taken and top-down technique is implement which reduce the complexity of algorithm as shown in the pie chart. The red portion shows the complexity of enhanced algorithm and blue portion shown the complexity of basic algorithm.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Frequent itemsets play very important role in our day to day life. Eclat algorithm is used to find frequent itemsets. But many problems are found in eclat algorithm while finding frequent itemsets. Like Large number of iterations is required for processing the items from huge database and also more escape time and more complexity has been found in eclat algorithm. To remove these problems, advanced eclat algorithm is developed which is based on calculating support values only. Number of iterations and escape time is get decreased by using transposed database. In the advanced eclat algorithm, Top-Down Approach is used by which complexity is decreased. As shown in the experimental results, proposed algorithm has high scalability and good speedup ratio. In this research, comparison had been made with calculating support values which shows that proposed algorithm is more efficient in terms of processing time and complexity and this algorithm has provide best way to find products.

In future, transposition of database will be applied on Apriori algorithm to analyse the performance in the term of escape time and number of iterations.

CHAPTER 8
REFERENCES

[1]SET-PSO-based approach for mining positive and negative association **rules**,Jitendra Agrawal, Shikha Agrawal, Ankita Singhai, Sanjeev Sharma, November 20140219-1377

[2] Mingjun Song, and Sanguthevar Rajasekaran**, "**A Transaction Mapping Algorithm for Frequent Itemsets Mining", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2014

[3] Moens, S.; Aksehirli, E.; Goethals, B., "Frequent Itemset Mining for Big Data," *Big Data, 2013 IEEE International Conference on* , vol., no., pp.111,118, 6-9 Oct. 2013

[4] "Frequent itemsets algorithms**"** International Journal of Machine Learning and Cybernetics, 2013, Page 1 Marghny H. Mohamed, Mohammed M. Darwieesh

[5] Kotiyal, Bina; Kumar, Ankit; Pant, Bhaskar; Goudar, R.H.; Chauhan, Shivali; Junee, Sonam, "User behavior analysis in web log through comparative study of Eclat and Apriori," *Intelligent Systems and Control (ISCO), 2013 7th International Conference on* , vol., no., pp.421,426, 4-5 Jan. 2013

[6] S Vijayarani and P Sathya, "Mining Frequent Item Sets over Data Streams using Eclat Algorithm" IJCA Proceedings on International Conference on Research Trends in Computer Technologies 2013 ICRTCT(4):27-31, February 2013

[7] Shaobo Shi; Yue Qi; Qin Wang, "FPGA Acceleration for Intersection Computation in Frequent Itemset Mining," *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2013 International Conference on* , vol., no., pp.514,519, 10-12 Oct. 2013

[8] Guo-Cheng Lan; Tzung-Pei Hong; Hong Yu Lee; Shyue-Liang Wang; Chun-Wei Tsai, "Enhancing the Efficiency in Mining Weighted Frequent Itemsets," *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* , vol., no., pp.1104,1108 13-16Oct.2013doi: 10.1109/SMC.2013.192

[9] Peng Jian; Wang Xiao-ling, "An improved association rule algorithm based on Itemset Matrix and Cluster Matrix," *Computer Science & Education (ICCSE), 2012 7th International Conference on* , vol., no., pp.834,837, 14-17 July 2012

[10] D. Gunaseelan, P. Uma, "An improved frequent pattern algorithm for mining association rules", International Journal of Information and Communication Technology Research Volume 2 No. 5, 2012

[11] Jaishree Singh, H. R. "Improving Efficiency of Apriori Algorithm Using Transaction Reduction ".International Journal of Scientific and Research Publications, 2011

[12] Chanchal Yadav, Shuliang Wang, Manjot Kumar,“An Approach to Improve Apriori Algorithm Based On Association Rule Mining”, IEEE,2011

[13] Kan Jin, "A new algorithm for discovering association rules," *Logistics Systems and Intelligent Management, 2010 International Conference on* , vol.3, no., pp.1594,1599, 9-10 Jan. 2010

[14] Noorhuzaimi, M.N.; Junaida, S.; Mazrul, R.M., "An analysis of network services using association rules," *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on* , vol., no., pp.469,473, 8-11 Aug. 2009

[15] Sang Lin; Hu-yan Cui; Ren Ying; Zhou-lin Lin, "Algorithm Research for Mining Maximal Frequent Itemsets Based on Item Constraints," *Information Science and Engineering (ISISE), 2009 Second International Symposium on* , vol., no., pp.629,633, 26-28 Dec. 2009

[16] Mahanti, Aniket, and Reda Alhajj. "Visual interface for online watching of frequent itemset generation in Apriori and Eclat." *Machine Learning and Applications, 2005. Proceedings. Fourth International Conference on*. IEEE, 2005.

Web-site References

1. http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Frequent_Pattern_Mining/The_Eclat_Algorithm
2. http://www.slideshare.net/wanaezwani/apriori-and-eclat-algorithm-in-association-rule-mining
3. http://www.borgelt.net/eclat.html
4. http://www.cs.umd.edu/projects/hpsl/classes/818s-s98/ppt/spring97-datamining/sld017.htm
5. http://ijcaonline.org/proceedings/icrtct/number4/10829-1048
6. http://en.wikipedia.org/wiki/Association_rule_learning
7. http://www.academia.edu/6435809/A_Survey_on_frequent_pattern_mining_methods-Apriori_Eclat_FP_growth
8. http://www.inside-r.org/packages/cran/arules/docs/eclat
9. http://stackoverflow.com/questions/8649759/generating-eclat-algorithm-implementation-from-apriori-algorithm
10. http://forum.ai-directory.com/read.php?5,68
11. http://www.philippe-fournier-viger.com/spmf/index.php?link=documentation.php
12. http://www.scientific.net/AMR.328-330.1896

Appendix:

FP: Frequent Pattern

CART: Classification and Regression Trees

CHAID: Chi Square Automatic Interaction Detection

LHS: Left Hand Side

RHS: Right Hand Side