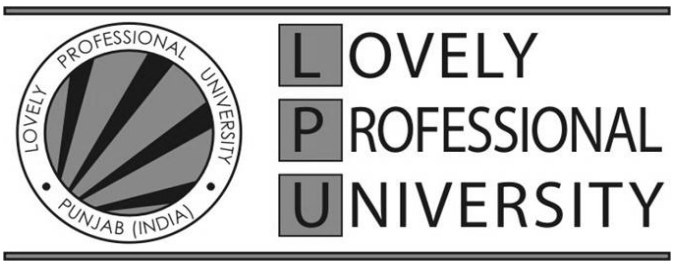


SPACS: Students' Performance Analysis and Counseling System using Fuzzy Logic and Association Rule Mining

A Dissertation

Proposal submitted

By

Ritu Banswal
Reg No. 11309157

To

Department of Computer Science and Engineering

In partial fulfillment of the Requirements for the
Award of the Degree of

Master of Technology in Computer Science and Technology

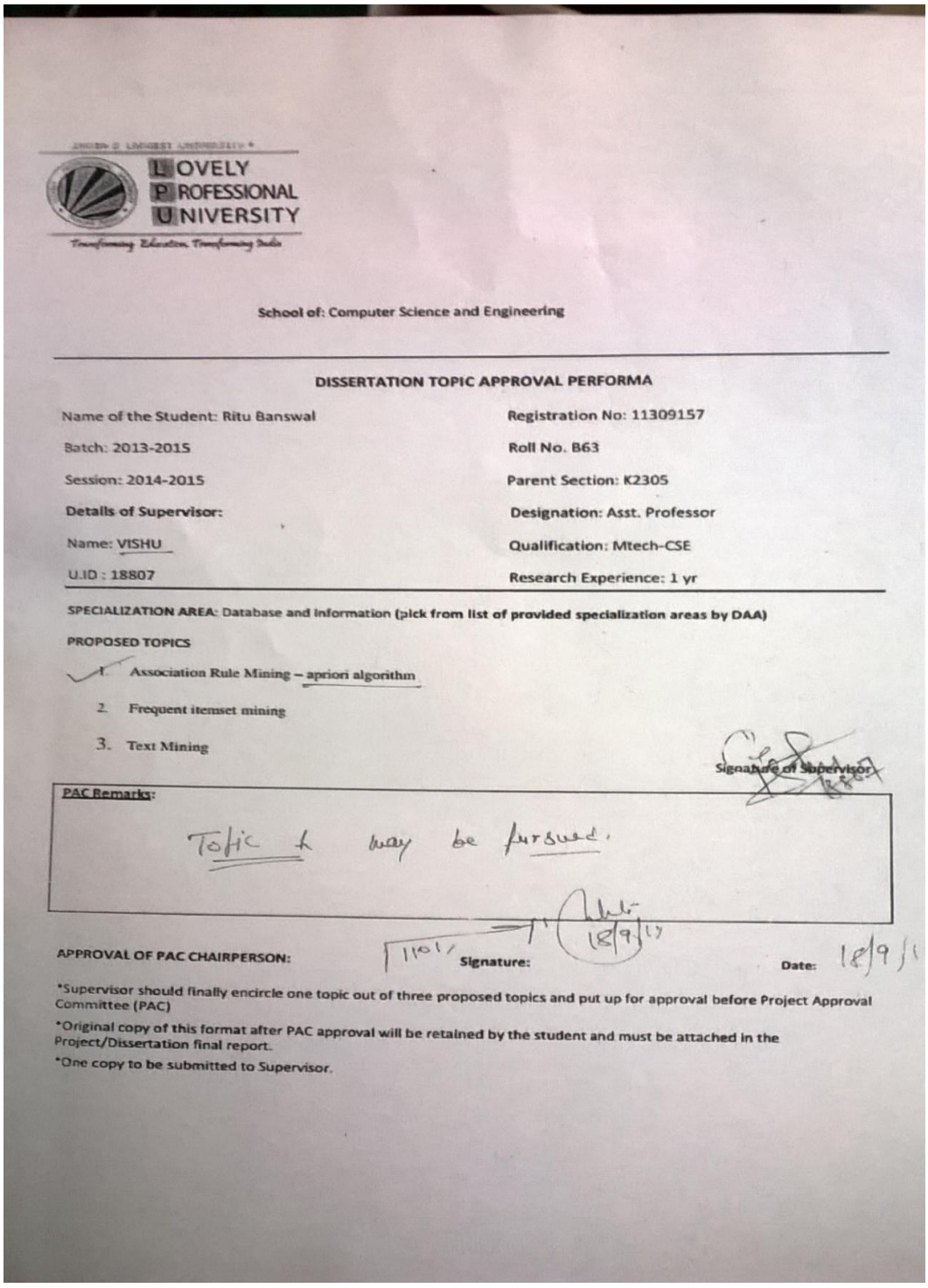
Under the guidance of

Ms.Vishu

Asst. Professor
Lovely Professional University, Jalandhar, Punjab

(May 2015)

PAC APPROVAL PAGE



LOVELY PROFESSIONAL UNIVERSITY

Transforming Education, Transforming India

School of: Computer Science and Engineering

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the Student: Ritu Banswal	Registration No: 11309157
Batch: 2013-2015	Roll No. B63
Session: 2014-2015	Parent Section: K2305
Details of Supervisor:	Designation: Asst. Professor
Name: VISHU	Qualification: Mtech-CSE
U.ID : 18807	Research Experience: 1 yr

SPECIALIZATION AREA: Database and Information (pick from list of provided specialization areas by DAA)

PROPOSED TOPICS

1. Association Rule Mining – apriori algorithm
2. Frequent itemset mining
3. Text Mining

Signature of Supervisor

PAC Remarks:

Topic 1 may be pursued.

APPROVAL OF PAC CHAIRPERSON: Signature: Date: 18/9/15

*Supervisor should finally encircle one topic out of three proposed topics and put up for 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

This study works on the use of data mining in expert system for taking decisions to solve real world problems and that proposed problem is providing quality education to students and enhances their personal growth for the society. The current study works on the existing students' data collecting within LPU, Punjab. After the data acquisition and analysis, data mining techniques and expert system approach are integrated to solve this issue. Student Performance Analysis and Counselling expert system is designed using association rules and fuzzy logic. Apriori algorithm will be applied for finding the critical factors that occurs frequently, affecting the student performance and based on that it generates strong association rules. This algorithm is the classical technique of ARM, works efficiently to mine the associations among the large dataset. In this study, MATLAB programming tool is used for implementing both mining and the fuzzy rule based expert system. Now this system would be able to analyze the performance of students either his/her academic performance increasing, decreasing or consistent. The complexity of the fuzzy expert system decreases by applying mining technique.

ACKNOWLEDGEMENT

I would like to present my deepest gratitude to Ms. Vishu for her support, guidance, advice, understanding and supervision throughout this dissertation study. I would also like to thank Mr. Prateek Agrawal and my previous mentor Mr.Arko Bagchi from LPU, Punjab and all the people who provided me with the facilities being required and conductive conditions for my M.Tech dissertation work.

I would also express my warm thanks for everyone who supported me throughout the course of M.Tech Dissertation. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the research.

DECLARATION

I hereby declare that the dissertation entitled, **SPACS: Student Performance Analysis and Counseling System using Fuzzy Logic and Association Rule Mining** 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: Ritu Banswal

Regn. No. - 11309157

CERTIFICATE

This is to certify that **Ritu Banswal** has completed M.Tech Dissertation titled **SPACS: Student Performance Analysis and Counseling System using Fuzzy logic and Association rule mining** 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 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 & Engineering.

Date: Signature of Advisor

Name

TABLE OF CONTENTS

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

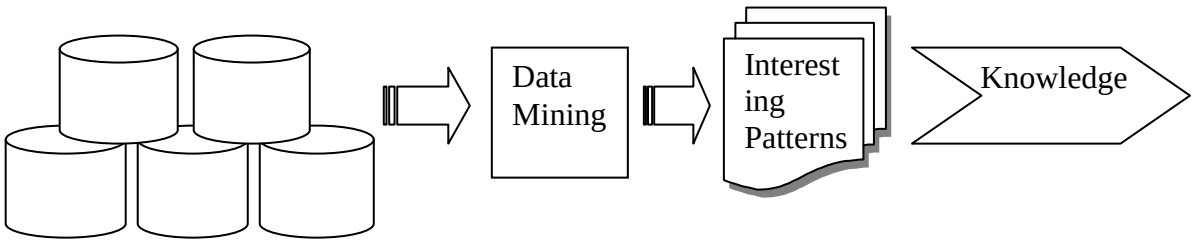
This chapter gives an introduction about the student performance analysis and counseling system that is designed using two different approaches that is fuzzy logic and association rule mining. The idea of the current research work comes from the previous researches on fuzzy association rule mining (FARM) [3]. FARM is a combined technique of data mining and artificial intelligence domain. Fuzzy logic concept comes under the artificial intelligence domain while association rule mining is one of the techniques of the data mining domain. In this, we study how these two domains interrelated to each other and how it is used in analyzing the student's performance.

FARM technique uses fuzzy logic to generate the interesting association rules. These hidden and interesting relationships help in taking the effective decision for the given problem. It is an enhancement to the classical association rule mining technique. The classical ARM uses the concept of crisp sets and has several drawbacks. Fuzzy ARM technique has applied on several applications but it is firstly used in the form of knowledge discovery in Fuzzy expert systems [3] [4] [2].

Here in our research work, we have applied one of the applications of Fuzzy ARM in analyzing the students' performance and used in developing the fuzzy expert system. But our methodology is different than the previous Fuzzy ARM technique and works with the aim of decreasing the complexity of the fuzzy expert system. As we know education plays an important role in one's life for personal growth but somehow there would arise some factors which deviates from his/her career path and affects the overall performance of the student [2]. The automated system would be required to analyze those factors and according to that an effective counseling could be provided. The complex task is to find the parameters which affect the overall performance of student. For solving this current problem, the integration of data mining and expert system results best approach. The components of each techniques used in the system along with their role have been illustrated with example.

1.1Data Mining

Data mining is a part of KDD (Knowledge Discovery in Databases) having vast research area used to analyze and extract the useful interesting patterns from the large datasets. It involves the concepts of machine learning, database systems, statistics and artificial intelligence [23]. The data mining have different alternative names like Knowledge discovery, data/pattern analysis, business analytics, etc. The knowledge acquired through the data mining technique can be applied into various applications like customer behavior, business management, market analysis, scientific exploration, knowledge based expert systems, etc. The objective of data mining is to extract or mine the hidden patterns from the data set and transform into the useful information called knowledge as shown in figure.1.1.



Large databases

Figure1.1. Data Mining- searching for interesting patterns knowledge in data.

As a knowledge discovery process, it involves data integration, data transformation, data cleaning, pattern evaluation, pattern discovery, data selection and knowledge presentation. The different steps when data mining viewed as knowledge discovery are as under:-

- **Requirement analysis:** The business problem must be clearly defined. On the basis this, the decision makers need to formulate the goals that data mining process have to be agreed upon.
- **Data selection:** This step includes the best source databases for the data that is required. Data warehouse stores the large amount of historical data. From this task relevant data is retrieved from the database.
- **Data cleaning:** This step is used to remove the noise and inconsistent data.
- **Data integration:** This step includes the data store where multiple data sources

may be combined.

- **Data transformation:** In this step data are transformed and load into appropriate forms for performing mining.
- **Data mining:** This is the essential step where intelligent methods are applied to extract the interesting patterns.
- **Pattern evaluation:** This step identifies the interesting pattern that comes from the data mining process and set to be considered as knowledge.
- **Knowledge Representation:** In this step the results are explained to the decision makers that visualizes and present the mined knowledge to the users.

Cross Industry Standard Process for Data Mining, known as **CRISP-DM** is a standard data mining process model that describes commonly used approaches of data mining to solve the problems [8]. CRISP-DM method breaks the data mining process into six major phases as shown in figure 1.2.

The six steps of process diagram are:-

1. Business Understanding
2. Data Understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

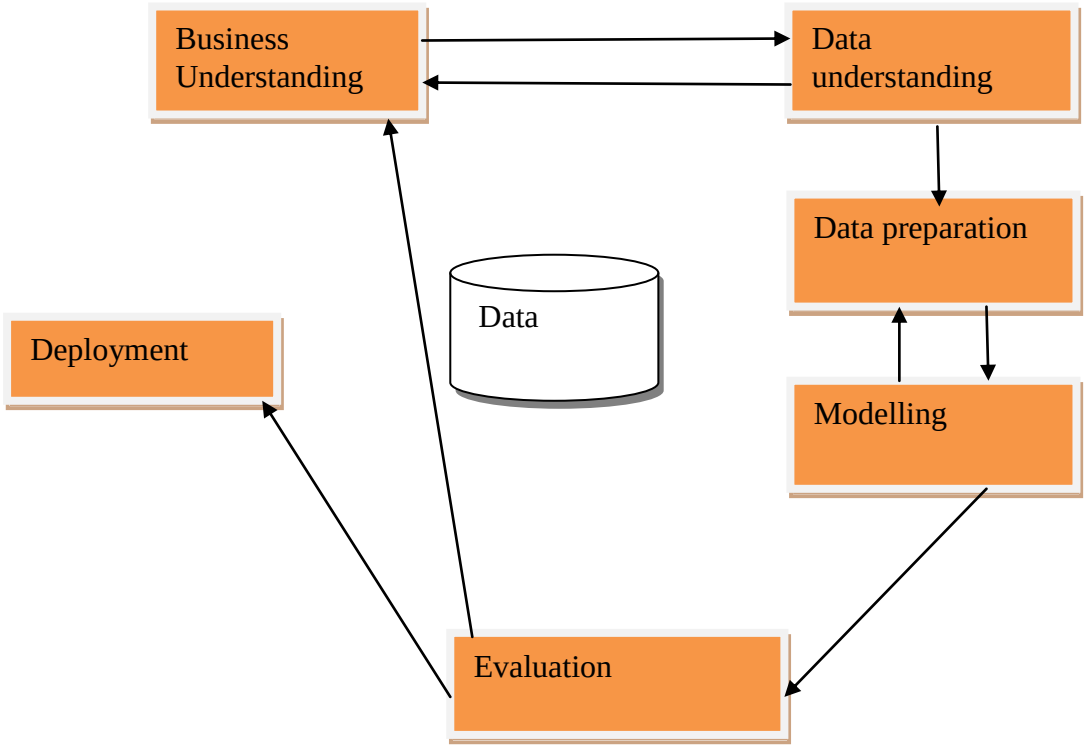


Figure 1.2 CRISP-DM process diagram

1.1.1 Data Mining Applications

Data mining is being widely used in variety of applications. It has incorporated many techniques from other domains such as statistics, machine learning, visualization, algorithms, information retrieval, pattern recognition, and many more. The success of the integration of data mining and its extensive applications provide dimensions to the data mining research and development. Business Intelligence and web search engines are the two highly successful and popular application examples of data mining.

The applications of data mining includes:-

- **Market segmentation:** It analyzes the characteristics of customer who buy the same product from the same organization.
- **Fraud Detection:** It identifies the different transaction that occurs with fraudulent.
- **Business Intelligence:** It deals with historical, current and statistical business information. Classification, prediction and clustering plays a central role in customer relationships and also analyzing the markets, supplies and sales.
- **Web search engines:** It deals with a specialized computer server that searches for the information on the web i.e. deals with the online data. The data mining

techniques varies from crawling to indexing and then move on to searching and at last search results can be personalized and made it context aware. It can also be termed as web data mining.

- **Data mining in education:** Data mining plays an important role in education learning with the applications of machine learning and statistics [24]. The aim is to mine the information about the learning process in order to make better decisions about the design and execution of learning environment. It comes under four phases like:-
    a) To discover the relationships in data using different techniques including classification, clustering, association rule mining, and sequential pattern mining.
    b) To check the validation of discovered data.
    c) Making predictions from the generated validations.
    d) Predictions are used for taking effective decisions.
- **Security and crime investigation:** Web mining are helpful in protecting the user system from cyber crimes such as cyber crimes, DOS attack, internet fraud, phishing attack, etc. Classification and clustering techniques are used in capturing the criminals [13] and have several applications.

1.1.2 Data Mining Techniques

Data mining includes different tools and techniques that are used to mine the useful knowledge from the large amounts of data. There are techniques that are currently used in business like association, classification, clustering, prediction and sequential pattern.
The data mining process specifies the two kinds of pattern approaches:-

1. Predictive
2. Descriptive

In Predictive data mining the output will be predicted based on the results generated for e.g. to diagnose a patient based on his/her medical test results. This approach is totally based on supervised learning method whereas unsupervised method is used in descriptive data mining. The descriptive approach generates the interesting patterns without defining the target for e.g. to evaluate the frequent patterns which have some relationships among the data items as in market basket analysis [1]. The actual task of data mining is to analyze the automatic or semi-automatic for the large quantities of data to extract the

unknown interesting patterns such as groups of data records called clustering, analyzing the unusual records called anomaly detection and finding dependencies called association rule mining [25].

1. **Frequent Pattern Mining (Descriptive)**

Frequent Pattern Mining is used to find the interesting patterns that occur frequently and have some relationships among the large no. of data items in a database. This learning method helps in analyzing the products that are frequently bought together. It is also known as market basket analysis for e.g. to find all the items that are frequently purchased with bread.

2. **Classification (Predictive)**

Classification Technique works with supervised learning. This is used to classify the different class attributes related to each other. The analysis is done with the help of training data which involves data objects whose class label is previously known for e.g. an email program that is used to classify whether the mail is spam or not and according to that it provides a solution.

3. **Clustering (Descriptive)**

Clustering Technique works with un-supervised learning. This involves data objects whose class label is previously not known. The analysis is done by dividing the data items into different clusters.

4. **Regression (Predictive)**

In this technique a function is defined called regression function which helps to find the relationship between the independent variables and the dependent variable.

1.1.3 Association Rule Mining

ARM (Association Rule Mining) comes under the descriptive approach used for frequent item set mining technique which helps to find the hidden patterns and relationships among those items in the datasets. It is a most widely used data mining technique in all real life applications like in business analysis, identifying the customer behavior, medical diagnosis, industry, etc. The most common example of association rule mining is “market basket analysis” to know the behavior of the customers by identifying which products is frequently purchased together [1]. Boolean variables are used to analyzing the buying patterns reflect items that are frequently associated or purchased together. The following

table shows the transactions for market basket analysis showing the items that customer brought together.

Table 1.1 Market basket Analysis-Transactions

Tid	milk	bread	jam	butter
1	0	1	1	0
2	1	1	0	0
3	1	0	1	0
4	1	1	1	1
5	1	1	1	0

This table shows a small database that contains the set of items ***I= { milk, bread, jam, butter}*** where (1 means presence and 0 means the absence of that item in a transaction). An example rule for the supermarket could be ***{bread, jam}→{milk}*** means that if the customer bought butter and bread together then, he also buy milk.

The association rule is defined as a state when X occurs then Y also occurs having confidence level ie.

$$X \rightarrow Y, \text{ where } X \cap Y \neq \emptyset .$$

Here X and Y may be single item or belongs to the item set. X is often referred as antecedent and Y as the consequent.

ARM is used to find the hidden and interesting item set that occurs frequently and have some relationships among the large datasets. Selecting the most interesting rules from all sets of possible rules depends on constraints like minimum thresholds on support and confidence.

For example, X and Y appear together in only 10% of transaction but when X appears there is 70% chance that Y also appears. This 10% of presence together is called support or known as prevalence of rule where as the other 80% of chance is called confidence or known as predictability of the rule.

- **Rule Evaluation Measures**:

  Support: It defines the percentage (s %) of transactions in D contains the union of both X and Y item sets.

$$\text{Support}(X \rightarrow Y) = P(X \cup Y)$$

Confidence: It defines the percentage (c %) of transactions in D having X that also contain Y item set.

$$\text{Confidence}(X \rightarrow Y) = P(X \mid Y)$$
$$= \text{support}(X \cup Y) / \text{support}(X)$$

- **Two step process:**
  1. To identify all the item sets that occurs frequently.
  2. From those frequent item sets generate the strong association rules.

E.g. A frequent item set {Chicken, Clothes, Milk} [support =3/7] and one rule from the frequent item set Clothes → Milk, Chicken [support = 3/7, confidence=3/3]

The strong association rules are generated being a complex task if they satisfy the minimum support threshold and minimum confidence threshold. This would helps for making decision for production of those products which is more beneficial for sell profit. A simple and classical algorithm known as *apriori* algorithm may be used to find the frequent item sets and then associations between them.

Apriori Algorithm is one of the classical ARM algorithms used to find the frequent item sets in large datasets for association rules. It uses the prior knowledge and follows the bottom up approach in which k item sets used to generate k+1 item sets.

The algorithm involves two concepts:-

- Find all frequent item sets that have minimum support.
- Based on frequent item sets, generating the strong association rules.

Works with two procedures:-

- *Join*: The candidate sets are generated by joining the frequent item sets in step-wise.
- *Prune:* Discard the frequent items if subset is not frequent.

Apriori Algorithm

Input: D, Database of transactions; min_support, minimum support threshold

Output: L, frequent itemsets in D

Method:

L1=find_frequent_1-itemsets (D);

```
for (k=2; Lk-1≠Φ; k++)
{
        Ck=apriori_gen (Lk-1, min_support);
   for each transaction t ∈ D
  {
        Ct=subset (Ck,t);
        for each candidate c ∈ Ct
        c.count++ ;
  }
Lk={ c ∈ Ck |c.count ≥ min_support }
}
 return L=Uk Lk ;

Procedure for apriori_gen(Lk-1: frequent(k-1)-itemsets)
for each itemset l1∈ Lk-1{
for each itemset l2∈ Lk-1{
if(l1 [1]= l2 [1])∧ (l1 [2]= l2 [2]) ∧…∧(l1 [k-2]=
l2 [k-2]) ∧(l1 [k-1]< l2 [k-1]) then {
 c= l1 ∞ l2;
if has_infrequent_subset(c, Lk-1) then
delete c;
else add c to Ck ;
}}}
return Ck;

Procedure for infrequent_subset(c: candidate k-itemset;Lk-1:frequent(k-1)-itemsets)
for each(k-1)-subset s of c
{
if s ∉ Lk-1 then
return true;
 }
return false;
```

Working of Apriori algorithm:-

1. Scan the database for identifying frequent items.
2. Candidate generation is done by joining process which helps to find the frequent item sets at different levels.
3. Pruning is done at each level if frequent item sets is not found and also does not satisfy the minimum support threshold.
4. Terminate the algorithm if there are no frequent item sets left.
5. Finally on the basis of resulted frequent item sets, strong association rules are generated which holds the condition that it satisfy the both minimum support and confidence threshold value.

1.2 Expert System

Expert system is an example of knowledge based system. These are the computer programs that are designed to provide the skills of an expert to non-experts. It is an automated system that helps to take effective decisions in certain circumstances as compared to human experts [2] [7]. The system is composed of two sub-systems: *inference engine* and *knowledge base.* Knowledge base represents the facts and rules, on the basis of these rules inference engine generates the results to the user.

The expert system consists of three main components:-

1. Knowledge base
2. Inference engine
3. User Interface

1.2.1 Elements of Expert System:-

The development of the expert system consists of following elements:-

- *Inference engine*- draws the conclusions deciding which rules are satisfied and prioritizing.
- *Agenda*- a prioritized list of rules, which are satisfied by facts or objects in working memory.
- *Knowledge Base*- information is stored in the form of rules.
- *Working Memory*- stores the global database of facts that are being used by rules.
- *Explanation Facility*- explains reasoning of expert system to user

- *Knowledge acquisition facility*- user acquires knowledge in an automatic way.
- *User interface*-– mechanism by which user and the system communicates.

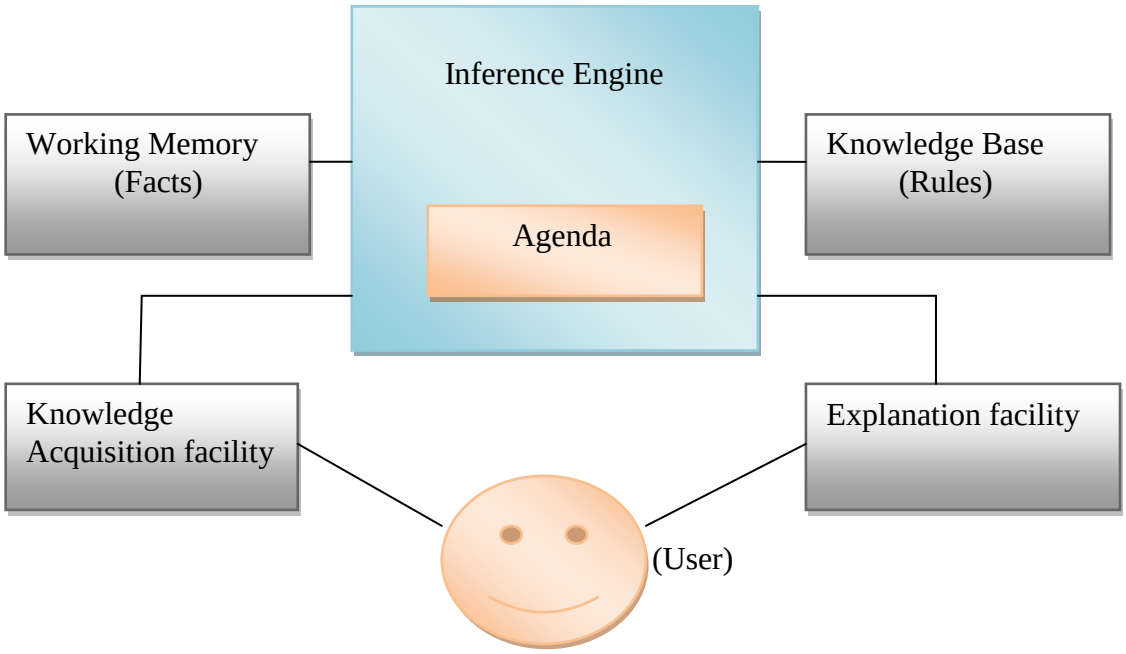


Figure1.3 Elements of Expert System

Inference Engine is a component and also called the brain of expert system in artificial intelligence. The engine applied the rules to the knowledge base and draws the conclusions deciding which rules are satisfied or having higher priority. There are two modes for working of inference engine: forward chaining and backward chaining. Forward chaining starts with the pre-defined facts and then asserts new facts where as backward chaining starts with the goals and then determine what facts must be asserted to achieve the goals.

Knowledge base is used to store the structured and unstructured information that is used by the system. Before storing the knowledge into the knowledge base, it must be acquired from domain experts or from other sources. This process of acquiring the knowledge is known as knowledge acquisition and it is done with the help of both technical skills and people skills on the part of knowledge engineer. The knowledge engineer then encodes and represents the expert's knowledge into knowledge base. There are different approaches for modeling the expert system: rule based expert system, frame based approach is used case-based system, and object oriented approach, etc. The rule based

expert system defines the knowledge base in the form of *if-then* rules. These systems are easy to formulate and decision making ability in a structured and modular way.

Knowledge acquisition subsystem is used to transfer and accumulates the problem solving expertise knowledge from domain experts or other sources to a computer program that is then used for constructing the knowledge base. Acquiring the knowledge from experts is a very difficult task in construction of expert system. Knowledge engineer helps to interact with the domain experts and then formulate the results according to the knowledge base.

User Interface provides a mechanism through which the user communicates with and commands the expert system through this subsystem.

1.2.2 How Expert System related to Data Mining

After extracting the information from data mining process, the discovered knowledge can cooperate with the domain expert system for generating an excellent knowledge base of an expert system [19]. An expert system is an automated computer system that takes decisions more accurately as compared to humans.

How it works?

The user interacts with this automated system with the help of inference engine. Inference engine is the heart of the expert system. It draws the conclusions from the facts and rules that defined under the knowledge base as shown in figure1.3. The inference engine in the expert system provides the facility to explain about the related query asked by the user and draws conclusions based on the expert domain knowledge stored in the knowledge base. Some applications of expert system are medical diagnose expert system [7], whether forecasting, prediction analysis, planning and scheduling, business decision making, etc.

Expert system solves the real world problem related to the different domain knowledge [12]. In the knowledge domain there can be possibility of uncertainty occurs. Most of the times data could not be remain consistent and we have to deal with uncertain data. Uncertainty occurs when rules are created it defines as when the things are not always true or false and also we are not having the complete knowledge about the particular domain. Fuzzy Logic is one of the techniques that deal with uncertainty problem.

1.2.3 Features of Expert System

- Ability to deal with uncertainty factors
- System would be easy to use and provides easy modification.
- System can be used with other system maintains transportability
- Explanation- provides reasoning process to reach certain goal
- The data related to particular problem is specifically specified
- Breaks the problem goal into sub-goals
- Inference solution from initial data

1.2.4 Fuzzy Logic

Fuzzy Logic is an expert system technique that deals with uncertain data [4]. The uncertain data defines in terms of 0 and 1. The truth value of the variables ranges in between either low, medium or high. Fuzzy logic uses probabilistic approach and mathematical approach while defining the truth values for particular problem. Imprecision and uncertainty are the two forms of fuzziness that may be in the aspects of measurement, probability, or descriptions. Imprecision is a kind of probability that deals with an uncertainty about the future occurrence of events or phenomena. For example a statement like "It might rain tomorrow" which shows a degree of randomness. This kind of randomness shows ambiguity and sometimes fuzzy in nature that is addressed by fuzzy logic.

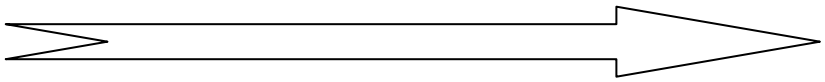
Two Concepts involved in Fuzzy Logic:

Linguistic Variables: are the variables whose values may either words or sentences in a natural and synthetic language.

Fuzzy If-Then Rule: in which the antecedent and consequent are the basic propositions that contain linguistic variables.

Fuzzy logic includes the concept of fuzzy sets. It is a set containing elements that can have partial degree of membership unlike in crisp sets, in which element either belongs to set or not.

For e.g a fuzzy logic for defining speeds as shown in figure1.4.



Slowest	Slow	Fast	Fastest
[0.1- 0.3]	[0.2- 0.6]	[0.5- 0.8]	[0.7- 0.10]

Figure1.4. Fuzzy Logic

In fuzzy logic, linguistic variables take the linguistic values that are associated with some degrees of membership in the fuzzy set. As defined from the figure 1.5 instead of a defining speed as a variable having a numerical value, it treated as a linguistic variable that assumes, linguistic values of *slowest* with a degree of membership ranges from[0.1-0.3], *slow* with a degree in between [0.2-0.6], etc.

Fuzzy set is useful for representing linguistic variables and a fuzzy set Z in X is defined by the function $\mu_z(x)$.

$$Z = \{x, \mu_z(x) \mid x \in X\}$$

where $\mu_z(x)$ is called the membership function of x in Z. This function maps each element of X to a membership value between 0 and 1. A membership function is a curve which shows the mapping of an input space to a membership value between 0 and 1 [ ] as shown in figure 1.5.

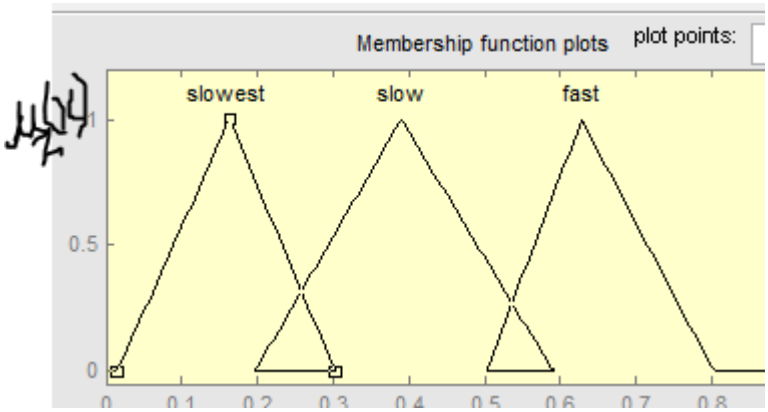


Figure 1.5 Fuzzy Membership Function

1.3 Fuzzy Expert System

Fuzzy expert system (FES) is an expert system that defines the membership functions and rules using fuzzy logic. The rules are defined in the form of IF-Then statements.
"*If value of A is low and value of B is high then C= Medium", where A and B works as input, C as an output*.

In this low, high and medium are the fuzzy membership function defined on A, B& C resp. The part of the rule "*if A is low and B is high"* is called the antecedent or premise that describes to what degree the rule applies, while the part of the rule "*then C is medium"* is called the consequent or conclusion that assigns a membership function to each of one or more output variables. The fuzzy sets define the membership functions to each one or more output variables. The set of rules defined in fuzzy expert system is stored in knowledge base according to the particular domain. These rules are in human readable form and they are easy to understand and interpret the results.
The fuzzy expert system is somewhat similar to traditional expert system but the difference lays in that it uses fuzzy set theory while inference the rules and known as fuzzy rule based expert system.

1.3.1 The development of fuzzy rule based expert system

The development of fuzzy rule based expert system consists of fuzzification, inference, knowledge base, defuzzification processes as shown in figure 1.6.

- **Fuzzification** is a process of transforming the crisp values into degree of membership functions for linguistic variables of fuzzy sets.
- **Fuzzy Knowledge Base** The Fuzzy system takes effective decisions with the help of this fuzzy knowledge base. The knowledge base consists of facts and rules about the particular domain. The rules are defined in the form of fuzzy if-then rules and the knowledge is being collected from domain experts.
- **Fuzzy Inference Engine** The Fuzzy expert system provides a user interface system with the help of inference engine. The facts are defined by the user and then inference engine provides expertise solution by applying some logic based on fuzzy knowledge base.
- **Defuzzification** is an inverse transformation process that converts the fuzzy sets into crisp set. The output obtained is a fuzzy set so to obtain a single output value,

defuzzification is needed. Different methods like centriod, max membership, bisector, SOM are used to defuzzify the fuzzy set.

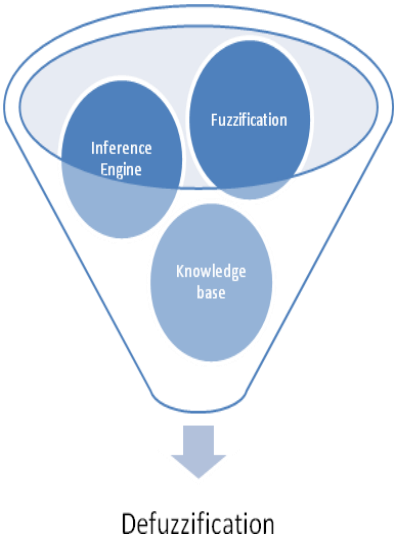


Figure 1.6 Fuzzy rule based expert system

1.3.2 Applications of fuzzy expert system

The fuzzy expert systems make use of fuzzy logic that deals with uncertainty and defines the membership functions for the linguistic variables define over the fuzzy set. The fuzzy expert system has certain applications in different fields. They are:-

1. Pattern recognition
2. Data analysis
3. Linear and Non-Linear control
4. Financial Systems
5. Operational research

1.3.3 Advantages of Fuzzy Expert System

- Works with complex system when its behavior not known what could be done.
- Works well when fast response is required.
- When no exact solution is required.
- Portable.
- Easy to understand the fuzzy rules.
- Uncertainty and explain ability.

CHAPTER 2

LITERATURE REVIEW

Prateek Agarwal *et.al* [2014] proposed Fuzzy rule based expert system which evaluates the overall performance of the students using fuzzy logic [2]. The data is collected within the university LPU, Punjab by one to one interaction with engineering students and faculties. After the discussion with higher authorities five critical factors came into the existence namely teaching factor, university system, and university environment, personal and family issues that affect the students' academic performance. Based on this information fuzzy expert system is developed which analyzing whether the performance will increase, decrease or remain consistent and on the basis of that it counsels the student for enhancing the overall growth. Identifying the most affecting parameters in the quality of academic performance is a very complex task. The expert system evaluates the performance based on various factors using fuzzy logic and thus the system is useful in improving the performance of the student.

Charanjit Bambrah *et.al* [2014] proposes the ARM to analyze the performance of the students and to enhance the quality of education in the educational institutions [5]. Apriori algorithm is used to extracts the set of rules, analyzes the given data to classify the student based on their performance in academics. Students are classified based on assignment, internal assessment, tests, attendance etc., for analyzing the performance of the student whether good, poor, or excellent. Then a report is generated to improve the result of student.

Neha Sharma *et.al* [2014] proposes an application of ARM and describes how the generated patterns are used by the experts for early detection and prevention of oral cancer [12]. Apriori algorithm is used to extract association rules among the clinical examination and history data with the help of data mining tool WEKA. The experiment results shows all the rules that have greatest confidence level thereby making useful for practitioners in early diagnose of oral cancer cavity and consequently helps in prevention of the disease.

N.Badal *et.al* [2014] proposed a new approach for implementation of apriori algorithm in MATLAB using Attribute Affinity Matrix [15]. Apriori algorithm is used to effectively mine the frequent data item sets from large database. The new methodology shows that the apriori algorithm when implemented in MATLAB performs better than the existing apriori algorithms. This approach reduces the execution time and is more reliable than classical algorithm.

Divya bansal *et.al* [2013] applied the Apriori algorithm on crimes concerning women database to discover and understand the underlying patterns involved in crime cases [13]. WEKA tool is used for analyzing the results and compared with other association rule algorithms.

K. Rajeswari Professor *et.al* [2013] designs a decision support system for medical diagnosis of patients and assists the doctors to predict risk of disease [7]. Medical data is used for mining frequent patterns using association rule algorithm, further genetic and fuzzy logic will optimize those generated rules to take effective decisions.

A. Rehab H. Alwa *et.al* [2013] proposed a new matrix approach to improve the apriori algorithm using MATLAB where the database transactions are saved. It avoids the repeated scanning of the database transaction where particular rows & columns are extracted and perform a function. The results can be easily visualized using graphical form display. In this, the algorithm prune columns of the matrix whose frequency count are less than the minimum support threshold and thus a new matrix are formed with item sets that satisfies the association rule. The new matrix formed consists of frequent item sets only and hence the matrix size reduces drastically. Thus the study concludes that the new matrix approach is faster and provides enhancement in terms of reducing computation time and memory space.

Olufunke O. Oladipupo *et.al* [2012] examines the students' performance ratings as against their pre-admission academic profile and relationship between the academic performances [3]. The study focuses to utilize fuzzy association rule mining technique in analyzing student profile that helps to take the admission according to his characteristics and the profile of the candidate. Fuzzy association rule mining (FARM) was used to analyze and identify the hidden relationships between the student's admission profile and

academic performance. This study helps to determine the academic profile of those students that are most admitted in the session and thus intimate the advisors to monitor the performance of student. The proposed approach uses the fuzzy logic for uncertain data set and then applies association rule mining to obtain and analyzing the results. The academic performance is evaluated using the Grade Point Average (GPA) for each particular session.

Dr. Varun Kumar *et.al* [2012] ARM technique is used to improve the quality of data by analyzing the affecting factors and thus to increase the success rates of students [11]. Different data mining methods like data pre processing, data cleaning, and data transformation are applied before preparing the data for mining the association rules. TANAGRA tool is used to mine association rules and analyzing the assessment of the students.

Hamid Eslami Nosratabadi *et.al* [2012], Fuzzy expert system designed to classify the customers of the banks using classification rules and association rules with the help of Apriori algorithm and CRISP-DM process, specifies the Credit Degree of banks' customers [8]. It combines the extracted rules of association mining and the knowledge of experts. The CRISP-DM (Cross-Industry Standard Process for Data Mining) model method is followed by this standard to study the 436 records of customers collected for the Saman Iranian Bank. It considers the four financial factors related to the customers, Current ratio, Debit ratio, Net Benefit Ratio, Claims collection period are inputs of database and the credit degree of customers as the output. The association rules are used to classify the data using Apriori algorithm. Then they concluded fuzzy expert system has been created based on these selected rules for evaluating the credit degree of Bank's customers.

Hamid Eslami Nosratabadi *et.al* [2011], proposes a fuzzy expert system that evaluates the association rules based on the interesting measures [9]. The system focuses on the case study of credit scoring for the bank customers. Interestingness measures are taken as input variables and interesting rule level is taken as output. FIS tool of MATLAB software has been used to develop the system. From this, resulted rules have been specified to rank the banking customers according to interesting nature of values. Apriori algorithm has been used for extracting the rules with 436 records based on four criteria.

Huiping Wang and Ruowu Zhong [2011] use Apriori algorithm- ARM technique to analyze the performance of the college students [10]. The data were processed, identifying the relations which affect the students' performance and then association rules were generated.

Neetesh Saxena *et.al* [2010] proposed to analyze the students' performance using fuzzy logic based on the two factors ie. Attendance and marks obtained on which the performance of the student mostly depends [4]. Fuzzy logic system is used to divide the data into various ranges- poor, very poor, good, fine, and excellent. Based on these data, membership functions are defined and rules are generated through which performance analysis is done and draws conclusions.

Dr. Ahmed Tariq Sadiq [2008] proposed study for premises of production rules. These rules are reduced for the rule based expert system using ARM [14]. Apriori algorithm is used for improving the complexity of search engine. The production rules are converted into transaction database and association rules are generated.

CHAPTER 3
PRESENT WORK

3.1 Scope of the study

Problem Formulation:

"SPACS: Students' Performance Analysis and Counseling Expert System using Fuzzy logic and Association Rule Mining".

As we know education plays an important role in one's life for the overall growth in the society. If the right education and right environment is not provided then it somewhat distracts the person from his/her path. There would be arising some factors like university environment, family issues, teaching factor, personal factor etc that affects the overall performance of the student. For solving this current problem, we designed an automated system which helps to analyze and identify all the critical factors that affects the performance of the student and then counsels him/her against those affecting parameters [2]. The integration of data mining and expert system results best approach our study we designed an expert system known as SPACS which works on the data sample of students collecting within the LPU, Punjab and makes use of fuzzy logic and association rule mining technique [1].

In 2014, Prateek Agarwal *et.al* [2] proposes the fuzzy rule based expert system that generates the 256 rules using the fuzzy logic for analyzing against the critical factors affecting the academic performance but this system is not fully efficient to generate the results. Its capabilities are not much enhanced using only fuzzy approach and lacks to optimize the results when including more factors. In 2012, Olufunke O. Oladipupo *et.al* examines the students' performance ratings as against their pre-admission academic profile and relationship between the academic performances [3]. The study focuses to utilize fuzzy association rule mining technique in analyzing student profile that helps to take the admission according to his characteristics and the profile of the candidate. In this first the data is defined under the fuzzy set then after that association mining is done to achieve the most interesting and frequently occurred items. The mining shows the best possible combinations between academic performances that help to take the decision against the admission process. But the system has the limitation that it does not generate the accurate results and the computation time for generalizing the fuzzy rules is more.

So for solving the limitations regarding previous studies, we have formulated the approach to design a fuzzy expert system that works under the results of the association rule mining technique. First the ARM approach is used to find the hidden patterns that having relationships among the datasets and generates the strong association rules. On the basis of that expert system would be able to analyze the performance of the student either increasing, decreasing, or remains consistent. This would help to reduce the complexity of the fuzzy expert system. It aims to provide the accurate results against the strong associativity between the critical factors and its affect on the student performance.

3.2 Objectives of the study

In this research we are proposed to design an expert system using fuzzy rules and association rules which would going to analyze the student's academic performance and counsels him/her according to those affecting parameters. The parameters could be like university environment, personal factor, family issues, teaching factor, academic grades etc. The objective of this study is to implement the system in any university or college for the overall grooming of the student.

The studies found that the ARM helps to find the strongest factors or rules which would definitely affecting the student's performance and on the basis of this fuzzy logic would be applied to develop a fuzzy expert system which helps to analyze and counsels the student's academic performance.

The proposed objectives are:-

1. To find the parameters that affects the students' academic performance.
2. Apply the association rule mining technique.
3. Fuzzy rules are generated based on the high prioritizing associating rules
4. To decrease the complexity of the fuzzy expert system.

3.3 Research Methodology

In this study, an expert system is created that works on the fuzzy rules from fuzzy sets and association rules extracting from association rule mining. The idea comes from the previous researches on FARM technique but in our present work, we do the reverse of FARM process and applied it on the application to analyze the student performance with the help of fuzzy expert system.

The process of research methodology for constructing the fuzzy expert system is shown in figure 3.1.

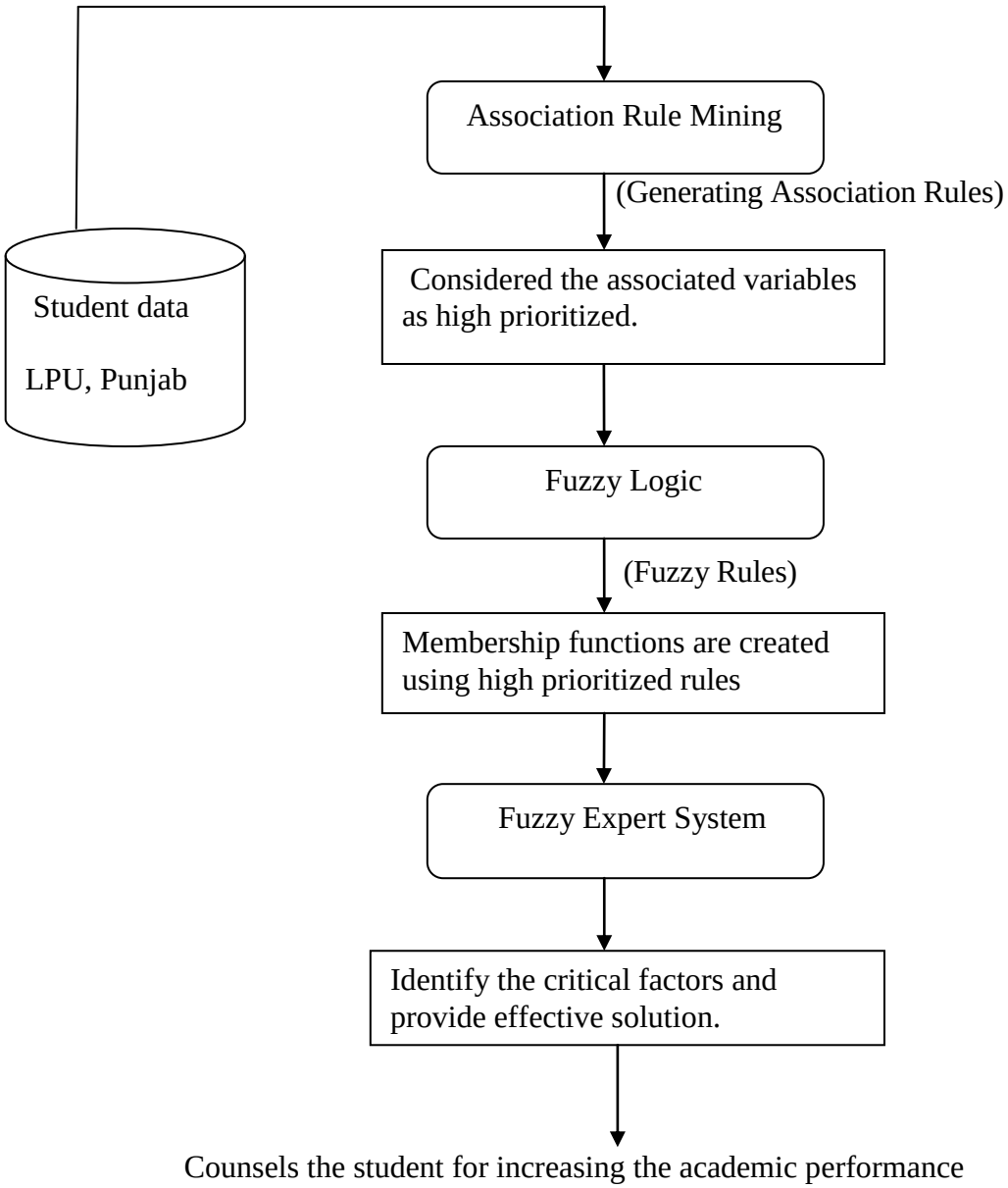


Figure3.1. Steps for Proposed research methodology

In this research, the steps for constructing such a fuzzy expert system from initial model to system evaluation are briefly defined as follows: -

Step 1: Clarify the objective.

Step 2: Data Collection

Step 3: Data Preparation

Step 4: Apply Association rule mining

Step 5: Fuzzy Expert System

Step1. Clarifying the Objective:

Once the problem has been clearly specified then the process of constructing the fuzzy expert system can begin. We need to first clarify the aim of the research work i.e. for what purpose we are going to construct the system.

Here, the main objective of our proposed work is to decrease the complexity and reduce the computation time of the existing work [2].The study focuses on the application that is going to analyze the academic performance of the students which would automatically tells whether the performance of the student will increase, decrease or remains consistent. According to these factors it would also counsels the student for his/her personal growth.

Step2.Data Collection:

Data collection is the very important step that is to be carried for the development of the proposed system. Once the objective is specified then we have to collect for solving the problem. The present study works on the existing student data that is collected within Lovely Professional University (LPU) Punjab by one to one interaction with engineering students and also a discussion was held with the highly experienced authorities of the university. The reason is to identify the critical factors that may affect the students' academic performance.

Five critical factors came into existence namely teaching methodology, personal factor, university environment, university system and family issues and they are further sub-divided into 36 factors [2]. Based upon the collected information a system, a system is developed that analyzes the performance of the student.

Step3: Data Preparation:

From the data collection step, we have collected the 569 records of the students for the student's performance analysis system. Since the researched dataset is not appropriate for

the next step i.e. for the association rule mining. We have to apply the data pre-processing or preparation process and makes it ready for the analysis. Data Pre-processing influences the results of the mining process in depth. It involves the noise removal, data reduction and makes it according to the relational database.

Here, we have considered the 36 factors related to the students that affects the performance includes confidence level, hesitation, university environment, teaching experience, teaching quality, distance among blocks, communication phobia, etc are treated as inputs for the database while performance factor as an output field. The 569 records of the transactional data are expressed by *yes* or *no* condition for the occurrence of the 36 critical factors represented as items in the dataset. Since the data preparation preserves the quantitative nature of the student data and defines them according to the item set for the association rule mining process.

Step4: Association Rule Mining

Association rule mining technique is now applied on the student dataset containing the 36 factors that are critical for the student academic performance. This technique is applied to extract the interesting and hidden patterns that occur frequently in the student dataset. It also extracts the dependencies between the interesting patterns and defines the relationships among the items in the dataset. Based on the dependencies, association rules have been generated and selected as the useful and proper rules.

Classical Apriori Algorithm is one of the ARM algorithms that is used for mining the frequent itemsets and generates the strong association rules from the student data which then considered as a high prioritized rules or factors affecting the student's performance. The algorithm is implemented in MATLAB programming tool for mining the frequent item sets and generating the association rules based on the those itemsets.

Implementation and Working of Association rule mining:-

Input:-

In this we have considered the 36 critical factors as data items and 569 records as the transactions for mining the associative rules from the student data set. This pre-processed data is stored on a separate .xls file for mining.

But for our proposed methodology we have converted this .xls file into text file that store in

the matrix form in MATLAB as shown in figure5.2. The student database is loaded into the MATLAB and it treated the dataset in the form of matrix with rows and columns. 569 transactions are considered as no. of rows while 36 critical factors are defined over columns as items. The data is represented in the form of Boolean values i.e. in 0 and 1. Here, the 0 means the absence of the item and 1 means the presence.

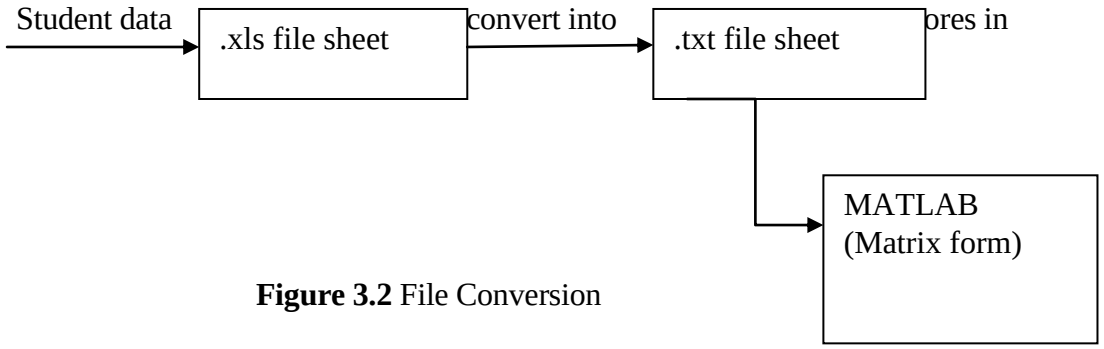


Figure 3.2 File Conversion

The input is taken as $m \times n$ matrix of binary transactions where m is the no. transactions and n is the no. attributes in the dataset. For example:

$$\begin{pmatrix} 1 & 1\,0\,0\,0\,1 \\ 2 & 1\,1\,0\,1\,0 \\ 3 & 0\,1\,1\,1\,0 \\ 4 & 0\,0\,1\,0\,0 \end{pmatrix}_{M \times N}$$

Minimum support threshold and minimum confidence threshold are also taken as input variables that are used to mine the frequent items in the dataset.

- *Minimum support threshold* is used to find the frequent item sets that satisfy the *support count.*
- *Minimum confidence threshold* is used to generate the association rules from those frequent item sets that satisfy the *confidence value.*

In our research, the algorithm works under the min_support of 0.1 and with 0.7 of min_confidence. Under these considerations the algorithm mines the frequent item sets that occur frequently in the dataset. The values are user defined and we have selected the average range for the minimum support and minimum confidence threshold. If the value of min_confidence kept greater than the 0.7 then we are not able to get the association

rules to check the possible relationships between the items.

Support: defines the percentage (s %) of transactions in D contains the union of both X and Y item sets.

$$\text{Support}(X \rightarrow Y) = P(X \cup Y)$$

The support count is calculated using the no. of attributes to the no. of transactions in the dataset. If the size of a particular item set is less than minimum support count then that item set is not consider as frequent itemset.

Confidence: defines the percentage (c %) of transactions in D having X that also contain Y item set.

$$\text{Confidence } (X \rightarrow Y) = P(X \mid Y)$$
$$= \text{support } (X \cup Y) / \text{ support } (X)$$

The confidence value is basically used to check the relationship of one item to the other item i.e. to find the hidden and interesting relationships among the dataset. Association rules are generated with the help of this confidence count.

- If the particular item set satisfies the confidence threshold then it is strongly associated with each other.
- If the item set satisfies both the threshold i.e. minimum support and confidence threshold then strong association rules are generated.
- If the item set does not satisfy the minimum support threshold then that itemset is pruned and also when no further candidate generation exist then at that point algorithm stops.
- Joining and Pruning are two processing elements of the apriori algorithm.

Output:-

Frequent itemsets are those item sets that occur frequently in the data set. The frequency of the itemset is based upon the satisfaction of the minimum support threshold. If it satisfies the support count then the item set is said to be frequent and consider for the generation of association rules.

In our work a cell array of frequent item sets of different sizes 1, 2,3,etc are generated with item set support >= min_support threshold. The sizes of frequent item set defines the item set consists of that particular items together. For example if frequent itemset is of size 2 i.e. freqItemsets {I1, I2} then it is shown that the candidate generation contains the elements of only 2 items.

Here, the frequency of item is found without scanning the database once again because the dataset is stored in the matrix form and it already exists. The results of frequent itemsets are stored in another matrix where as the association rules are stored in the separate text file.

Rules are defined under the cell array of *2×1* where the first cell depicts the left side of the rule i.e. the antecedent part where as the second cell depicts the right side of the rule i.e. the consequent part.
For example, if first rule shows $\{1, 2\} \rightarrow \{3\}$ then the cell array stores the rules in the form

Rules {1}{1}=[1, 2] and Rules{2}{1}=[3]

The purpose of implementing the apriori algorithm in MATLAB is that it treated each and every item in the matrix form which is easier to manipulate. Matrix approach is an enhancement to apriori algorithm that the computation time and memory space reduces [15]. MATLAB is a high-level language and provides an interactive environment for numerical computation, visualization, and programming [29]. Using this tool you can analyze data, develop algorithms, and create models and applications. It allows matrix manipulations, plotting of functions and data, implementation of different algorithms, creation of user interface application. The tool can be used in any of field like engineering, science and economics. It also supports graphics and graphical user interface programming.
There are various data mining software tools for mining. Association rule mining is a built in algorithm in some of the tools like in TANAGRA [27] and WEKA [28]. They work under the java programming language but the limitation of using these tools is that they takes lot of time to generate the results and also does multiple scanning of the database. When the same classical apriori algorithm is implemented in MATLAB, it scans the database only once. Hence reduces the computation time and memory space.

The results generated in MATLAB are easier to formulate and understand.

Step5: Fuzzy Expert System

This step involves the designing of the fuzzy expert system for student performance analysis. The process involves the different phases.

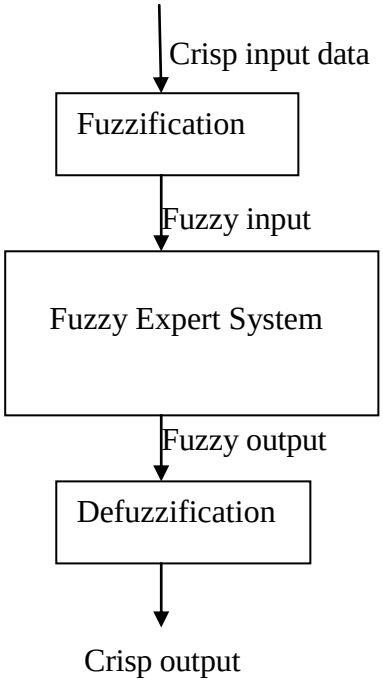


Figure3.3. Fuzzy inference system for analyzing students' performance

I. Fuzzification

Here the results of association rule mining are categorized according to the strong association rules extracted from the student dataset. Fuzzification process is then done to transform the crisp values into degree of membership functions for linguistic variables of fuzzy sets. Input and output variables are defined based on the mining process.

The factors that occurs frequently in the mining process and through which the association rules are generated are selected as input variables for fuzzy sets. In our study, we have extracted the 17 factors as input like communication phobia, confidence level, and hesitation, bad company, financial issues etc from the 36 factors. Now these 17 factors are categorized into five major categories as shown in table3.1. These major categories are taken as inputs for the fuzzy inference system and performance factor as output.

Input variables (Major factors)	**Sub-factors**
Personal factor	*communication phobia, confidence level, hesitation, communication phobia, confidence level, hesitation, bad company*
Family factor	*financial issues*
University factor	*distance among blocks, university atmosphere, overcrowded strength of students, campus disturbances*
Academic factor	*placement policy, examination pattern, academic schemes*
Teaching factor	*Teaching quality, teaching experience, teaching style.*

Table3.1. Categorization of input variables

Assign the input and output variables

In this system 5 input variables are defined as:-

- *Personal factor*
- *Family factor*
- *University factor*
- *Academic factor*
- *Teaching factor*

The output variable is:-

- *Performance factor*

Once the input and output variables are defined then fuzzy sets will be specified using the concept of fuzzy logic. The variables are fuzzified with the help of defined membership functions.

II. Fuzzy sets

After selecting the input and output variables the next step is to define fuzzy sets and their membership functions. In this system, for each input factor three fuzzy sets are defined i.e. low, medium, high and for output five fuzzy sets are defined i.e. inconsistent increase, inconsistent decrease, consistent increase, consistent increase, and consistent.

Triangular shaped built-in membership function is used to plot the fuzzy sets is described as “trimf” function. It is triangular in shape and is used as by default built-in membership

function.

Syntax

y = trimf(x, parameters)

y = trimf(x, [p q r])

Where p, q, r are the membership range for fuzzy sets and based on that graph is plotted.

For example: trimf(mf2, [3 5 9])

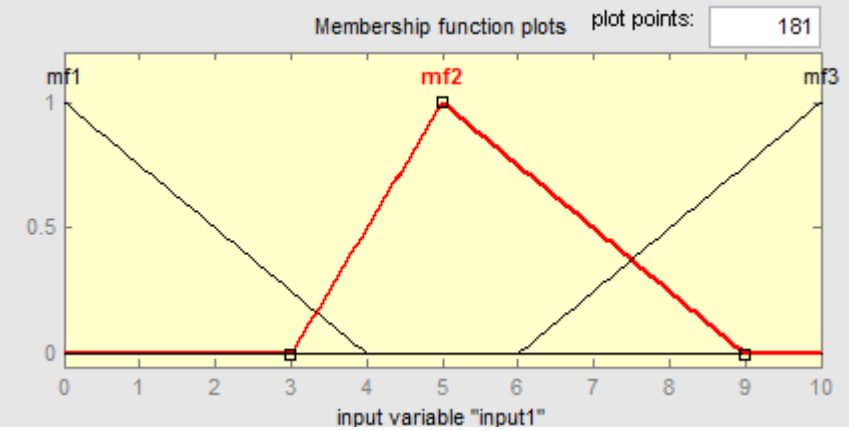


Figure 3.4 Membership curve for trimf, mf2= [3 5 9]

The values of the fuzzy sets for each input and output variables are as shown in the table 3.2.

Factors	**Membership range for fuzzy sets**				
Input Variable	**Low**		**Medium**		**High**
	0-30		20-60		50-100
Output Variable	**Consistent Decrease**	**Inconsistent Decrease**	**Consistent**	**Inconsistent Increase**	**Consistent Increase**
	0-30	20-50	40-60	50-80	70-100

Table3.2 Fuzzy linguistic variables and their membership values

III. Fuzzy rules

After defining the membership ranges for the input and output linguistic variables, a knowledge base is created by defining the fuzzy rules. The knowledge base of rule based fuzzy expert system stores knowledge in the form of fuzzy rules and the system draws

conclusion with the help of these rules.

The fuzzy rules are in the form of *If-Then* statements. These *if-then* rule statements are used to formulate the conditional statements that are a part of fuzzy logic.

If: *condition1* and *condition 2* and *condition3* and *condition 4*

Then: take *action 3*

For example

"*If value of A is low and value of B is high then C is Medium", where A and B works as input, C as an output*.

In this system, 204 rules are generated. Some of the rules are as follows:-

- If (personal_factor is low) and (family_factor is low) and (university_factor is low) and (academic_factor is high) and (teaching_factor is high) then (performance is inconsistent decrease) (1)
- If (personal_factor is medium) and (family_factor is medium) and (university_factor is medium) and (academic_factor is high) and (teaching_factor is high) then (performance is inconsistent decrease) (1)
- If (personal_factor is high) and (family_factor is low) and (university_factor is high) and (academic_factor is low) and (teaching_factor is low) then (performance is conistent_decrease) (1)
- If (personal_factor is medium) and (family_factor is high) and (university_factor is medium) and (academic_factor is medium) and (teaching_factor is high) then (performance is consistent) (1)
- If (personal_factor is high) and (family_factor is low) and (university_factor is high) and (academic_factor is low) and (teaching_factor is high) then (performance is consistent) (1)
- If (personal_factor is high) and (family_factor is medium) and (university_factor is high) and (academic_factor is medium) and (teaching_factor is high) then (performance is consistent) (1)
- If (personal_factor is high) and (family_factor is medium) and (university_factor is medium) and (academic_factor is high) and (teaching_factor is medium) then (performance is inconsistent decrease) (1)

- If (personal_factor is low) and (family_factor is medium) and (university_factor is medium) and (academic_factor is medium) and (teaching_factor is low) then (performance is inconsistent decrease) (1)
- If (personal_factor is low) and (family_factor is high) and (university_factor is high) and (academic_factor is high) and (teaching_factor is low) then (performance is inconsistent decrease) (1)
- If (personal_factor is medium) and (family_factor is medium) and (university_factor is low) and (academic_factor is low) and (teaching_factor is medium) then (performance is consistent_increase) (1)
- If (personal_factor is low) and (family_factor is high) and (university_factor is medium) and (academic_factor is low) and (teaching_factor is medium) then (performance is inconsistent increase) (1)
- If (personal_factor is high) and (family_factor is high) and (university_factor is medium) and (academic_factor is low) and (teaching_factor is medium) then (performance is inconsistent increase) (1)

These fuzzy rules are implemented on the basis of the 26 rules that are extracted from the association rule mining with 70% confidence threshold among the 569 records of dataset with 36 factors. After dividing the 17 factors into 5 major categories, we reunite the associations between the items and generate the possible combinations with each other. For example:-

Communication phobia → teaching quality

Overcrowded strength of students, academic schemes → teaching quality

Hesitation, overcrowded strength of students, teaching quality → teaching experience

They are defined as like:-

a. 1→5
b. 3, 4 →5 => 3→5,
   4→5
   3→4→5
c. 5, 3→5

As 3→5 combination is previously occur then we discard the rule c. We take the maximal possible combinations of each rule.

IV. Defuzzification

In this step defuzzification of the fuzzy output values are done. The output value obtained from previous step is in the fuzzy form and thus to get into single crisp value, defuzzification process is needed. Different methods like centriod, max membership, bisector, SOM are used to defuzzify the fuzzy set. But in our study, we use centriod method for defuzzify the output.

CHAPTER 4
RESULTS AND DISCUSSIONS

The present study works on the existing student data that is collected within Lovely Professional University (LPU) Punjab by one to one interaction with engineering students and also a discussion was held with the highly experienced authorities of the university.

The results are shown in figure4.1.

M27 ▾ fx Quality Level of Teaching, Too Much Strictness / Leniency of faculty Member

A	B	C	D	E	F	G	H	I	J	K	L	M
Student Name	Reg. No.	Pragramme & Batch Year	Class XII Percentage	Pass Out Board / Institute	Class X Percentage	TGPA	CGPA	1. Personal Factors	2. Family Factors	3. University Environment	4. University Sysytem	5. Teaching
Raj Kumar	10904070	B.Tech (H) CSE ; 2009	59	Punjab S	48	3.5	3.5	Communication Language Fobia, Low Level of Confidence, Hesitation to take initiatives, Physical or Medical Fitness	Financial Issues in Family, Over Expectations of Parents from their ward, Lack of awareness of family members about the education	Distance among various Blocks	Placement Related activities and policy, Examination Pattern	Too Much Strictness / Leniency of faculty Member
Pranjal	10809082	B.Tech (H) CSE ; 2008	70	CBSE Ranikhet	84	4	3.79	Opting the program or trade by chance., Intellectual Limitations, Bad Company or Poor Friend Circle, Hesitation to take initiatives, Wastage of Time because of Daily Transportation, Negative Attitude of student (Filled by the counselor at the end of interview)		Cosmopolitan Environment, Weather and Atmosphere in University, Inconvenience with the infrastructure of Classrooms		Quality Level of Teaching, Teaching Experience of Faculty Members, Too Much Strictness / Leniency of faculty Member, Student-Faculty Association, Age difference between Teacher and Student
								Wastage of Time because of Daily Transportation, Physical or Medical Fitness, Negative		Cosmopolitan Environment, Weather	Academic factors related to schemes and curriculum,	Teaching Style, Quality Level of Teaching, Teaching Experience of Faculty Members, Too Much Strictness / Leniency

Figure 4.1 Data collection within LPU, Punjab

We have collected the 568 records of the students for the student's performance analysis system and considered the 36 factors related to the students that affects the performance includes confidence level, hesitation, university environment, teaching experience, teaching quality, distance among blocks, communication phobia, etc. Since the data preparation preserves the quantitative nature of the student data and defines them according to the item set for the association rule mining process.

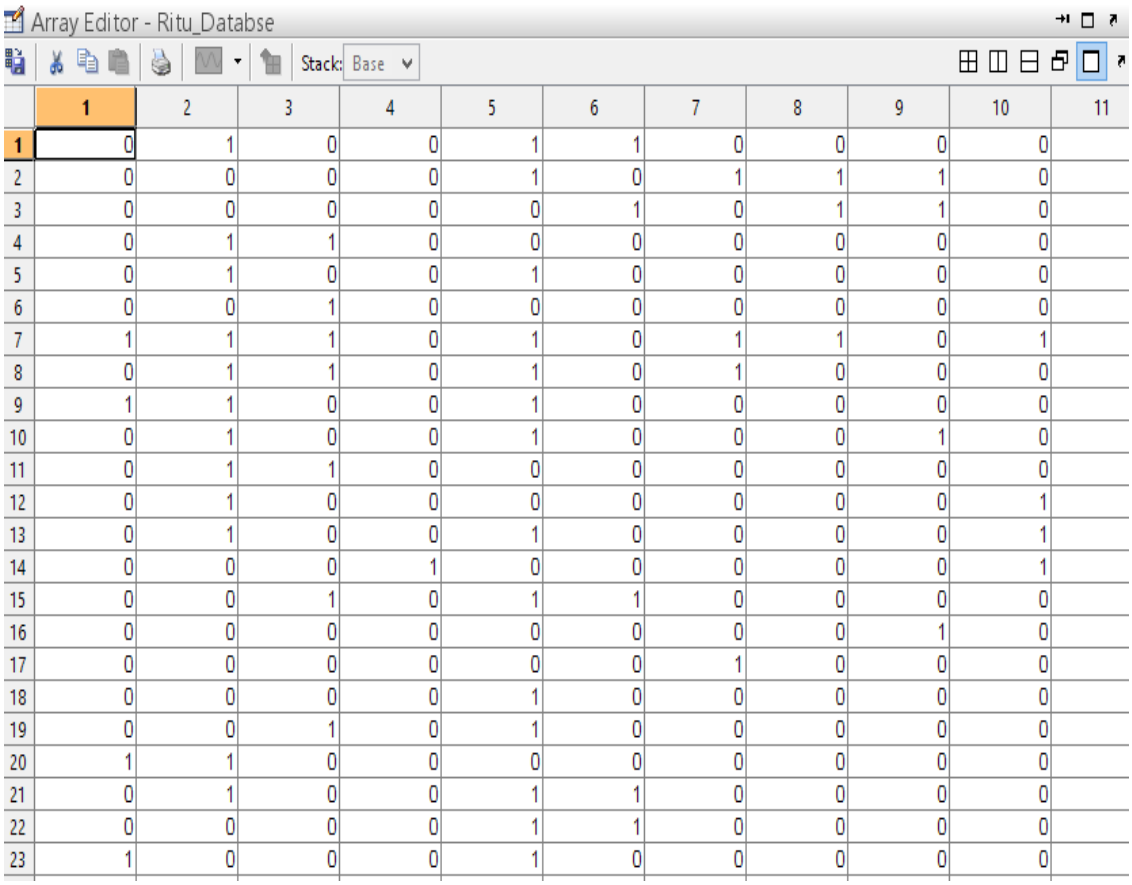
The results are shown in figure 4.2

B	C	D	E	F	G	H	I	J	K
Communication Language Fobia	Confidence level	Home sickness	Involvements in extra curricular activities	Hesitation	Medical Fitness	Intellectual Limitations	Negative Attitude	Time wastage due to transportation	Mutu Bond
yes	yes	no	no	yes	yes	no	no	no	no
no	no	no	no	yes	no	yes	yes	yes	no
no	no	no	no	no	yes	no	yes	yes	no
no	yes	yes	no	no	no	no	no	no	no
no	yes	no	no	yes	no	no	no	no	no
no	no	yes	no	no	no	no	no	no	no
yes	yes	yes	no	yes	no	yes	yes	no	yes
no	yes	yes	no	yes	no	yes	no	no	no
yes	yes	no	no	yes	no	no	no	no	no
no	yes	no	no	yes	no	no	no	yes	no
no	yes	yes	no	no	no	no	no	no	no
no	yes	no	no	no	no	no	no	no	yes
no	yes	no	no	yes	no	no	no	no	yes
no	no	no	yes	no	no	no	no	no	yes
no	no	yes	no	yes	yes	no	no	no	no
no	no	no	no	no	no	no	no	yes	no

Figure 4.2 Data Pre-processing

The student database is loaded into the MATLAB and it treated the dataset in the form of matrix with rows and columns. 569 transactions are considered as no. of rows while 36 critical factors are defined over columns as items. The data is represented in the form of Boolean values i.e. in 0 and 1. Here, the 0 means the absence of the item and 1 means the presence.

The results are shown in figure 4.3.



	1	2	3	4	5	6	7	8	9	10	11
1	0	1	0	0	1	1	0	0	0	0	
2	0	0	0	0	1	0	1	1	1	0	
3	0	0	0	0	0	1	0	1	1	0	
4	0	1	1	0	0	0	0	0	0	0	
5	0	1	0	0	1	0	0	0	0	0	
6	0	0	1	0	0	0	0	0	0	0	
7	1	1	1	0	1	0	1	1	0	1	
8	0	1	1	0	1	0	1	0	0	0	
9	1	1	0	0	1	0	0	0	0	0	
10	0	1	0	0	1	0	0	0	1	0	
11	0	1	1	0	0	0	0	0	0	0	
12	0	1	0	0	0	0	0	0	0	1	
13	0	1	0	0	1	0	0	0	0	1	
14	0	0	0	1	0	0	0	0	0	1	
15	0	0	1	0	1	1	0	0	0	0	
16	0	0	0	0	0	0	0	0	1	0	
17	0	0	0	0	0	0	1	0	0	0	
18	0	0	0	0	1	0	0	0	0	0	
19	0	0	1	0	1	0	0	0	0	0	
20	1	1	0	0	0	0	0	0	0	0	
21	0	1	0	0	1	1	0	0	0	0	
22	0	0	0	0	1	1	0	0	0	0	
23	1	0	0	0	1	0	0	0	0	0	

Figure 4.3 Load the data in MATLAB

By applying the association rule mining technique, frequent itemsets and association rules are generated using apriori algorithm implemented in MATLAB tool. The hidden and interesting patterns among the student dataset are shown in the figure 4.4. Frequent itemsets are generated by joining and pruning procedures of the algorithm. For each itemsets, different dimensions are specified in the matrix form.

```
FreqItemsets =

    [30x1 double]    [85x2 double]    [59x3 double]    [4x4 double]

See the file named Ritu_DatabseRules.txt for the association rules

Freq_set1 =

     1
     2
     3
     5
     6
     7
     9
    10
    11
    12
    13
    14
    15
```

Figure 4.4 Frequent item sets with their dimensions

From these interesting and frequent patterns, association rules are generated and are stored in the text file. The rules are categorized under the higher confidence value that is specified earlier in the program. The rules are as shown in the figure 4.5.

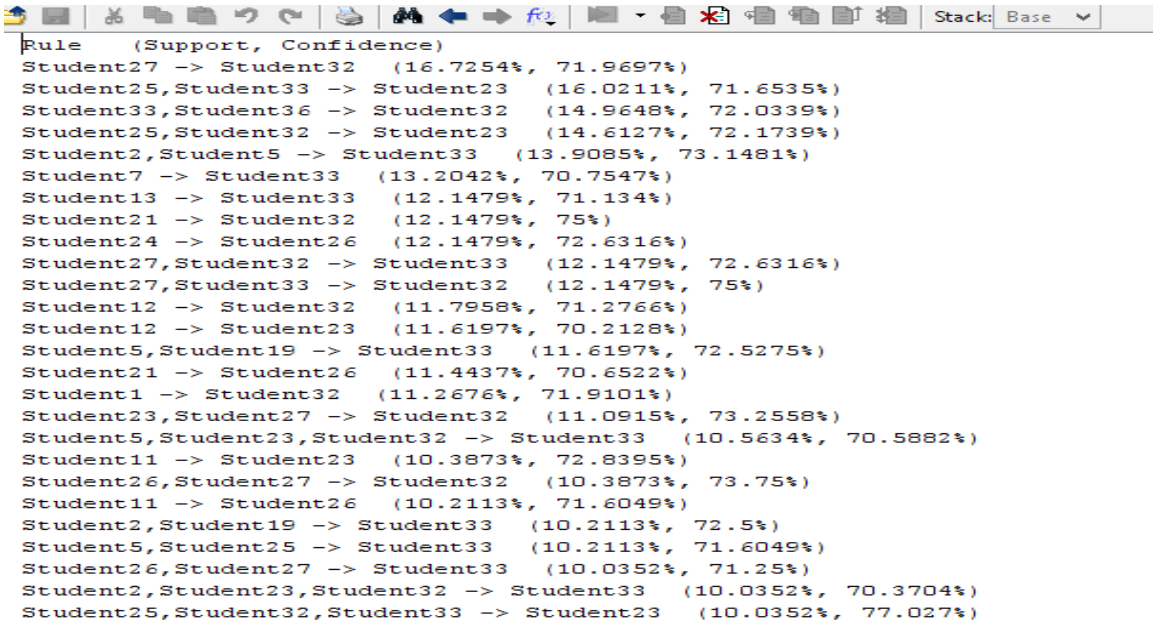


Figure 4.5 Association rules generated from student dataset

17 factors are extracted as hidden and interestingness that occurs frequently in the dataset and have some associations among the data items. The association rules are generated based on the minimum support count and minimum confidence value. Now these factors are categorized into five major categories i.e. personal factor, family factor, university factor, academic factor and teaching factor. The calculation of output remains transparent to the user and is based on fuzzy logic. In our study, Mamdani's fuzzy inference engine is used.

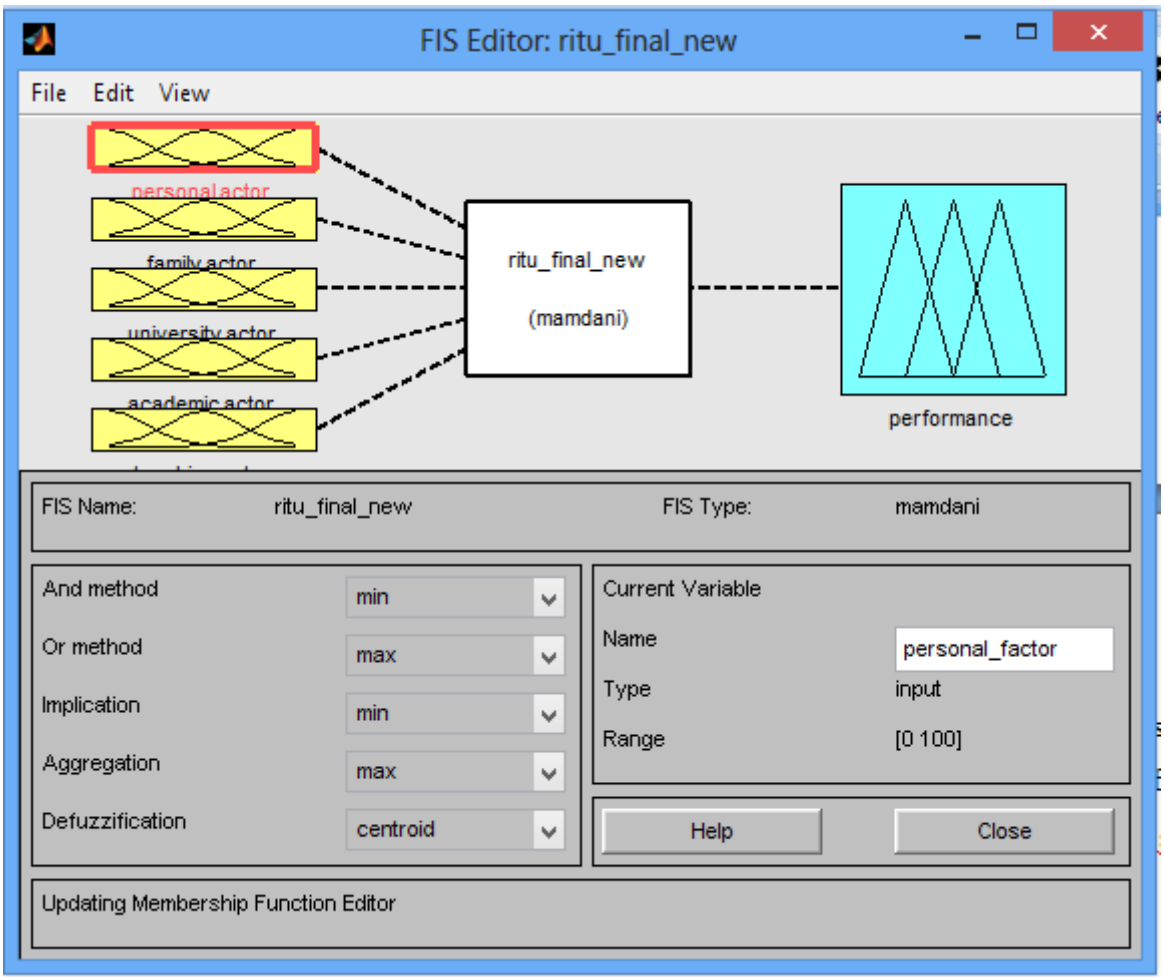


Figure 4.6 Fuzzy Inference System

Membership range of fuzzy set for each input variable and output variable are shown as:-

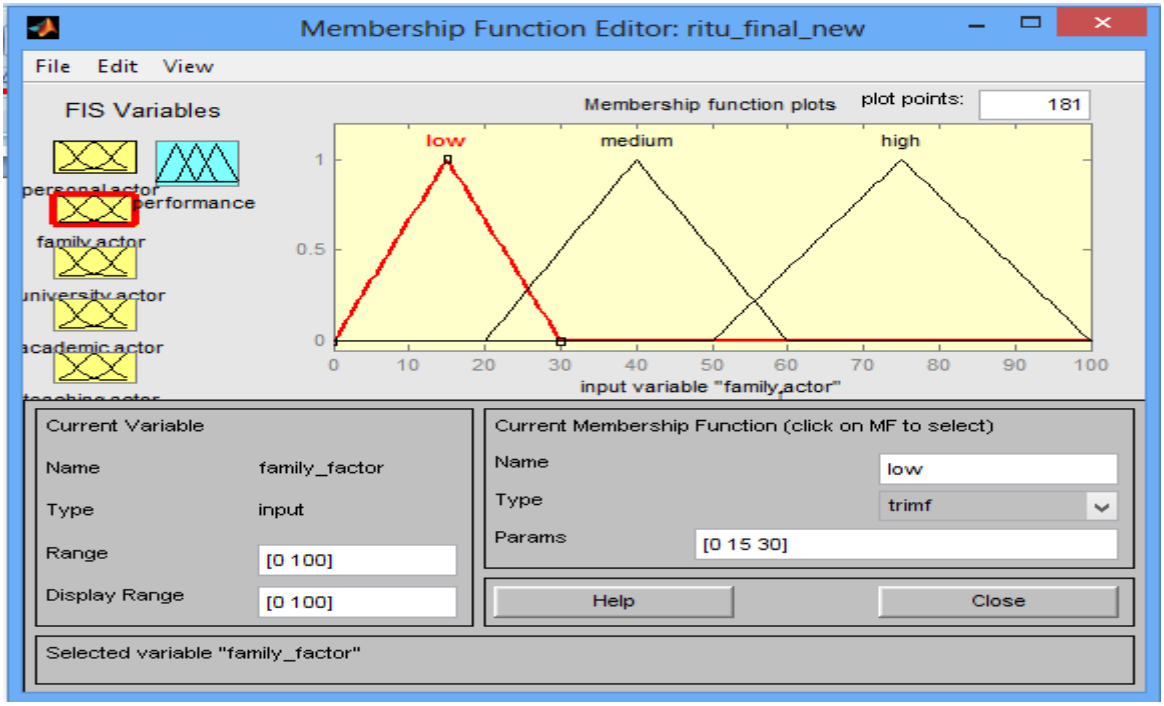


Figure 4.7 (a) Fuzzy Membership range for input variable

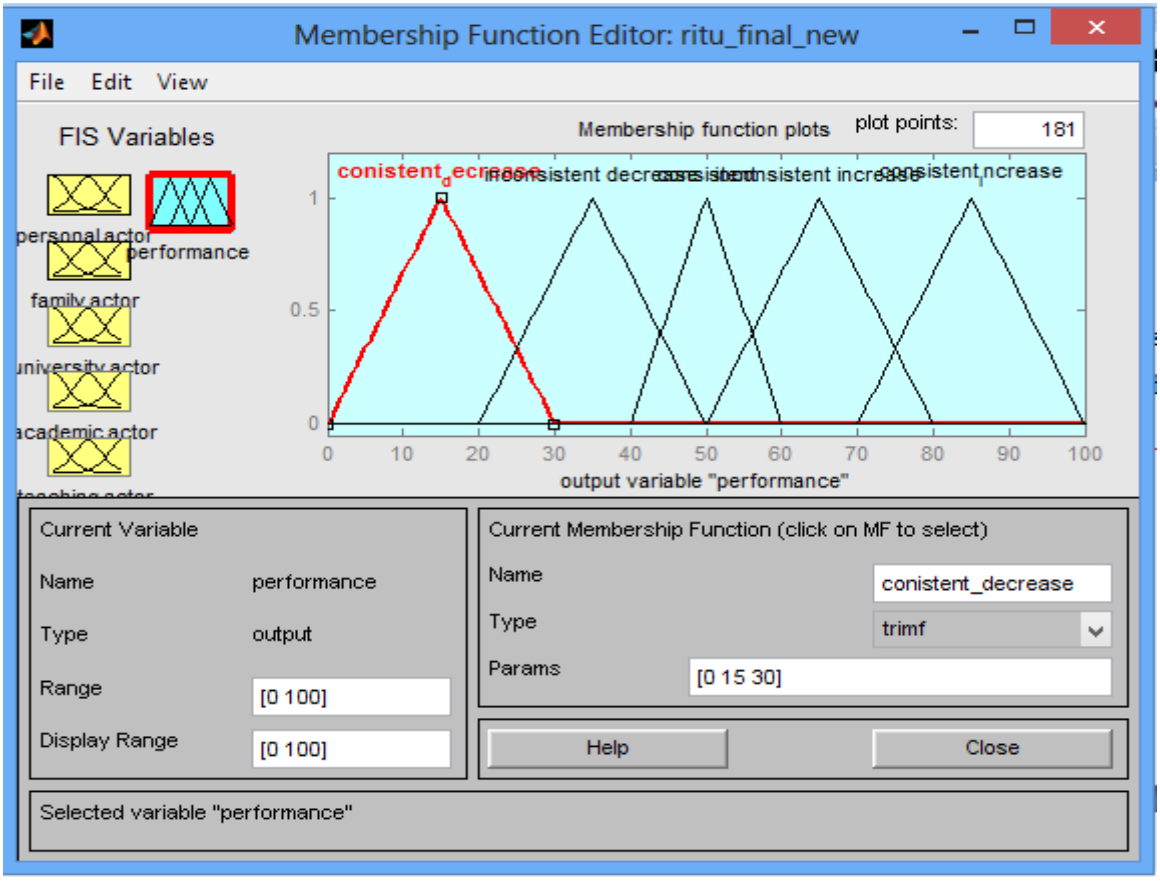


Figure 4.7(b) Fuzzy membership range for output variable

In this system, 204 fuzzy rules are generated based on the results of association rules. The results are shown in figure4.8. Rule editor is used to generate the fuzzy rules based on the results of the rules extracted from the association rule mining.

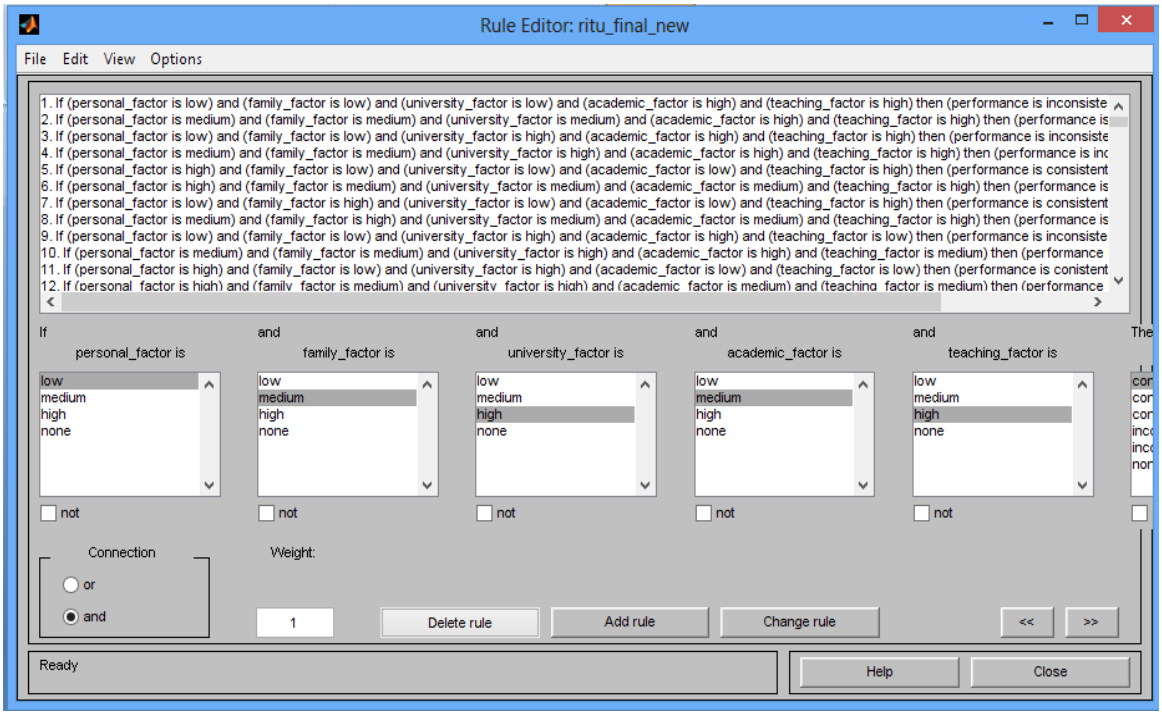


Figure 4.8 Rule editor

Rules are viewed under the rule editor that helps to edit the values of each input as to check the possible output.

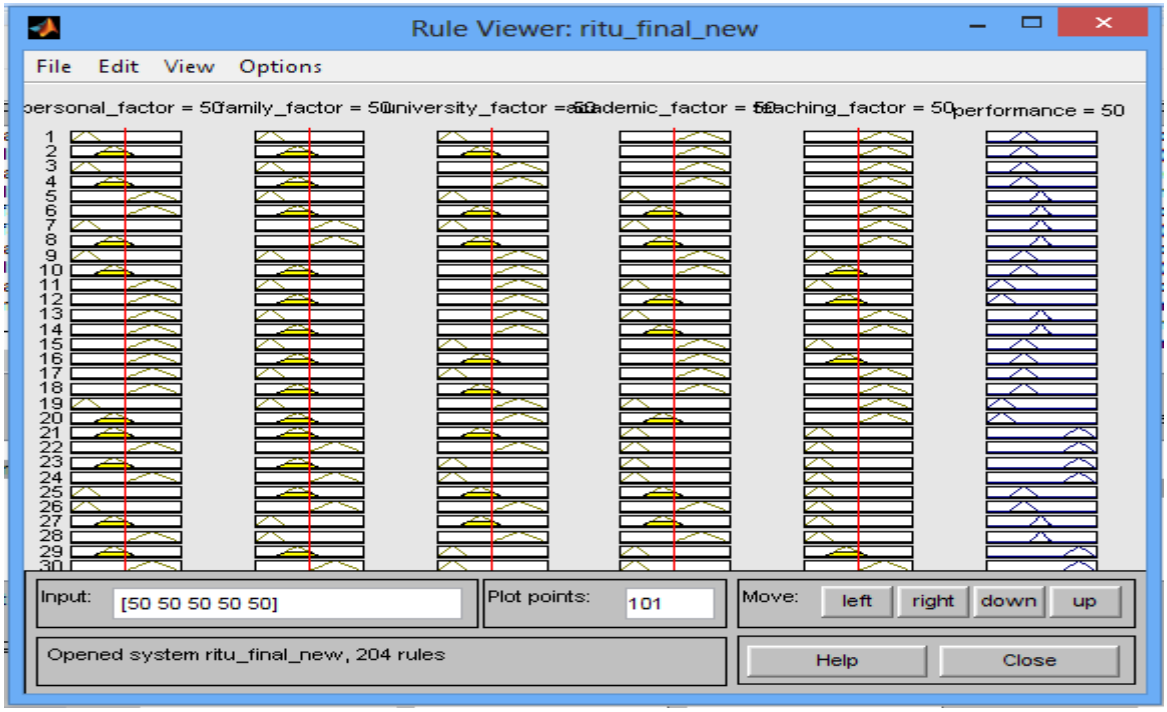


Figure 4.9 Rule Viewer

Surface viewer is used to view in three dimensions. The 3-D curve represents the mapping from two inputs and one output.

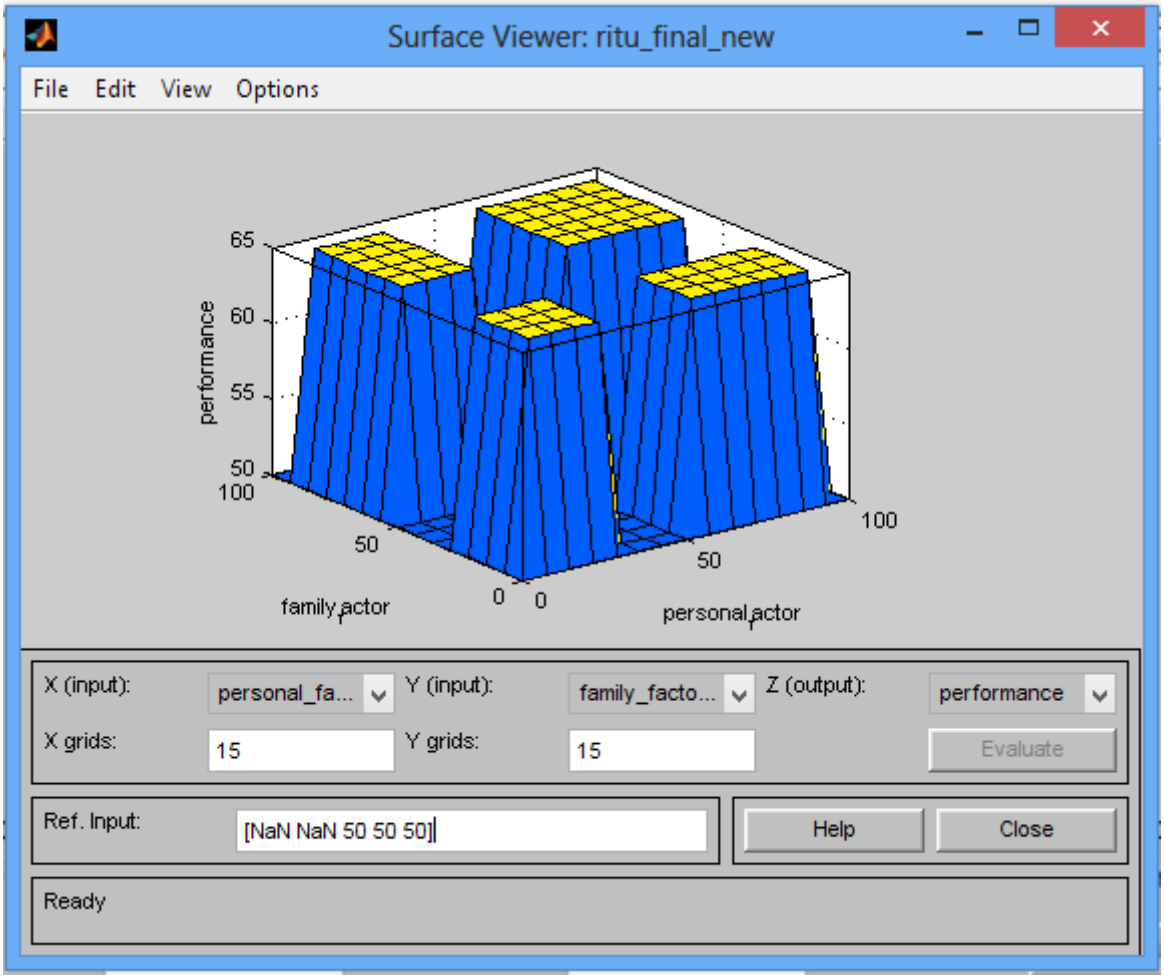


Figure 4.10 Surface Viewer

Hence, this students' performance analysis system using fuzzy logic and association rule mining generates the 204 rules for analyzing the performance of the student either inconsistent increase, inconsistent decrease, consistent, consistent increase, or consistent decrease. Prateek Agarwal *et.al* [2] proposes the fuzzy rule based expert system that generates the 243 rules using the fuzzy logic for analyzing against the critical factors affecting the academic performance. This proves that the current system with new approach using association rule mining and fuzzy logic decreases the complexity and also draws the accurate conclusions by the fuzzy inference engine.

Testing is also done by using different defuzzification techniques. This system is defuzzified for the methods like centroid, SOM, LOM, MOM and bisector to test the robustness of the rules and knowledge engineering. The test shows that the system performs better for all methods but not on SOM method because the ranking does not match with the other results as shown in table 4.1.

Sno	Personal factor	Family factor	University factor	Academic factor	Teaching factor	Defuzzification using different techniques								
						centroid		Lom		Mom		Som		bisector	
						values	ranks	Values	Rank	values	rank	values	Rank	values	rank
S1	75	0	30	45	65	50	4	50	4	50	4	50	5	50	4
S2	90	15	45	25	75	58.6	5	56	5	50	5	44	4	57	5
S3	79	7	65	48	8	15	1	23	1	15	1	7	2	15	1
S4	60	18	90	30	20	15	2	24	2	15	2	6	1	15	2
S5	18	39	30	83	91	35	3	44	3	35	3	26	3	35	3
S6	0	45	50	90	60	50	4	50	4	50	4	50	5	50	4

Table4.1. Testing of results

CHAPTER 5
CONCLUSIONS AND FUTURE SCOPE

5.1 Conclusion

The system is developed that helps to analyze the students' academic performance against the critical factors considered while collecting the information. It works under the two approaches one is association rule mining technique and the other is fuzzy logic. At last, fuzzy rule based expert system is designed that evaluates the performance of the student either it is inconsistently increasing, inconsistent decrease, consistent increase, consistent decrease or remains consistent. In this, the previously existing student dataset is used that is collected within the LPU, Punjab by one-to-one interaction to students and discussion with higher authorities specifying the critical factors affecting the student academic performance. First, the association rule mining approach is applied to generate the strong association rules from the frequent itemsets that occurs frequently in the dataset. Then fuzzy logic is used to define the fuzzy rules for the inference engine based on the results of the mining process. An association rule describes the strong associativity among the critical items that are hidden and interesting patterns in the dataset. The associativity depends on the higher confidence threshold that is specified by the user.

Hence the study concludes that by using association rule mining technique, the complexity of the fuzzy rule based expert system decreases. The accurate results are generated by the system and thus reduce the computation time for the analysis of the students' performance. The testing shows that it works better for all defuzzification methods except SOM method.

5.2 Future scope

The scope of the system can be extended by including more generalised association rules that are extracted based on the higher confidence value specified by the user. Besides using the classical apriori algorithm, we can extend the system to use the improved mining algorithm. The system can also be improved by considering the more hidden and interesting factors that may affect the performance of the student.

CHAPTER 7
REFERENCES

I. Research Papers

[1] Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." In *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207-216. ACM, 1993.

[2] Kaur, Parwinder, Prateek Agrawal, Sanjay Kumar Singh, and Leena Jain. "Fuzzy rule based students' performance analysis expert system." In *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on*, pp. 100-105. IEEE, 2014.

[3] Oladipupo, Olufunke O., Olanrewaju J. Oyelade, and Dada O. Aborisade. "Application of Fuzzy Association Rule Mining for Analysing Students Academic Performance." *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, Issue 6 No. 3, 2012.

[4] Saxena, Neetesh, and Kajal Kaushal Saxena. "Fuzzy Logic Based Students Performance Analysis Model for Educational Institutions." *VIVECHAN International Journal of Research* 1 (2010): 79-86.

[5] Bambrah, Charanjit, Minakshi Bhandari, Nirali Maniar, and Vandana Munde. "Mining Association Rules in Student Assessment Data." In *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, Vol. 3, Issue 3, 2014.

[6] Wang, Hua, Ping Liu, and Hongyang Li. "Application of improved association rule algorithm in the courses management." In *Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on*, pp. 804-807. IEEE, 2014.

[7] K.Rajeswari, Mahadev Shindalkar, Nikhil Thorawade, Pranay Bhandari. "DSS Using Apriori Algorithm, Genetic Algorithm and Fuzzy Logic" *International Journal of Engineering Research and Applications* (IJERA), Vol. 3, Issue 4, 2013, pp.132-136

[8] Nosratabadi, Hamid Eslami, Ahmad Nadali, and Sanaz Pourdarab. "Credit Assessment of Bank Customers by a Fuzzy Expert System Based on Rules Extracted from Association Rules." *International Journal of Machine Learning and Computing*, Vol. 2, No. 5, (2012)

[9] Nosratabadi, Hamid Eslami, Ahmad Nadali, and Sanaz Pourdarab." "Evaluating Discovered Rules from Association Rules Mining Based on Interestingness Measures Using Fuzzy Expert System" *International Journal of Machine Learning and Computing (IJMLC)*, Vol. 2, No. 5, IEEE 2011

[10] Wang, Huiping, and Ruowu Zhong. "Application of Association Rules in Analysis of College Students' Performance." In *Applied Informatics and Communication*, pp. 402-408. Springer Berlin Heidelberg, 2011.

[11] Kumar, Varun, and Anupama Chadha. "Mining association rules in student's assessment data." *International Journal of Computer Science Issues* 9, no. 5 (2012): 211-216.

[12] Sharma, Neha and Hari Om. "Significant patterns for oral cancer detection: association rule on clinical examination and history data." *Network Modeling Analysis in Health Informatics and Bioinformatics* 3, no. 1 (2014): 1-13, Springer

[13] Bansal, Divya, and Lekha Bhambhu. "Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women." *International Journal Of Advanced Research in Computer Science and Software Engineering, ISSN* 2277 (2013) vol 3, issue 9, pp. 54-62.

[14] Sadik, Ahmed T. "Premises reduction of rule based expert systems using association rules technique." *Int. J. SofiComput* 3, no. 3 (2008): 195-200.

[15] N. Baddal, S. Bagga, "Implementation of Apriori Algorithm in MATLAB using Attribute Affinity Matrix" IJARCSSE, vol. 4, issue 1, jan 2013.

II. Books

[16] Jiawei *et.al*, "Mining Frequent Patterns, Associations and Correlations," *Data Mining: Concepts and Techniques*, San Francisco, CA, itd: Morgan Kaufmann 5, 2012

[17] Negoita, Constantin Virgil. *Expert systems and fuzzy systems*. Benjamin-Cummings Publishing Co., Inc., 1984.

[18] D. A. Waterman *et.al*, *Building expert system*, Addison-Wesley Pub., 1983, p. 405-420 1983.

III. Reports and Documents

[19] Pirjo Moen, Association rules [Online] Available FTP: www.cs.helsinki.fi Directory: u/ronkaine/tilome/luentomateriaali File:TiLoMe-140305.pdf

[20] SJ Cunningham, Use Data Mining to support Expert System [online] Available FTP: www.cs.waikato.ac.nz Directory: ml/publications/1993 File: Holmes-Cunningham-DM93.pdf

[21]AnitaWasilewska,AprioriAlgorithm[online]AvailableFTP:www3.cs.stonybrook.eduDirectory:cse634/lecture_notes File:07apriori.pdf

[22] Oladipupo, Olufunke Oyejoke. "A Fuzzy Association Rule Mining Expert-Driven Approach to Knowledge Acquisition." PhD diss., Covenant University, 2012.

IV. Websites

[23] http://en.wikipedia.org/wiki/Expert_systems

[24] http://en.wikipedia.org/wiki/Data_mining

[25] http://en.wikipedia.org/wiki/Educational_data_mining

[26] http://en.wikipedia.org/wiki/Association_rule_learning

[27] Ricco Rakotomalala, "TANAGRA: un logiciel gratuit pour l'enseignement et la recherche", in Actes de EGC'2005, RNTI-E-3, vol. 2, pp.697-702, 2005.

[28] http://www.cs.waikato.ac.nz/ml/weka/

[29] http://in.mathworks.com/products/matlab/?nocookie=true

CHAPTER 7
APPENDIX

LIST OF ABBREVIATIONS

ARM- Association Rule Mining

CRISP-DM- Cross-Industry Standard Process for Data Mining

FARM- Fuzzy Association Rule Mining

KDD- Knowledge Discovery for Databases

LPU- Lovely Professional University

MATLAB- Matrix Laboratory

WEKA- Waikato Environment for Knowledge Analysis

SOM- Small of maximum

LOM- Large of maximum

MOM- Medium of maximum