



# Analysis and Predication of performance student using classifier filtering data Mining technique

Rohitansh Singh Thakur<sup>#1</sup>, Dr. Sandeep Dubey<sup>\*2</sup>,

<sup>#1</sup> Scholar Student, <sup>#2</sup> Asst. prof., CSE Department, RKDF University  
Bhopal, (M.P.), India

<sup>1</sup>singh.rohitansh786@gmail.com (Size 9)

<sup>2</sup>sandeep.rkdfgroup@gmail.com

## Abstract

The aim of this thesis paper is to analyse student performance using data mining. Data mining techniques are widely used in educational Sector to find new concealed patterns from student's data. The concealed patterns that are exposed can be used to understand the problem arise in the educational sector. To understand student performance, a survey was conducted by RKDF University Bhopal with the support from the College and the peer learners of different branches. In present Academic organization scenario are facing many challenges like data sparsity, collection of data, scalability, synonymy, normalized, validated and revalidated etc. In direction to proceed further through data mining based on the understanding of the survey, data was collected form the central database of RKDF University where the main aim was to relate CGPA and student performance. We investigated different properties of the data; collected and developed a classification hypothesis in instruction to apply data mining algorithms. In this research a machine learning tool called WEKA. In the future further rigorous study to match between demographic data

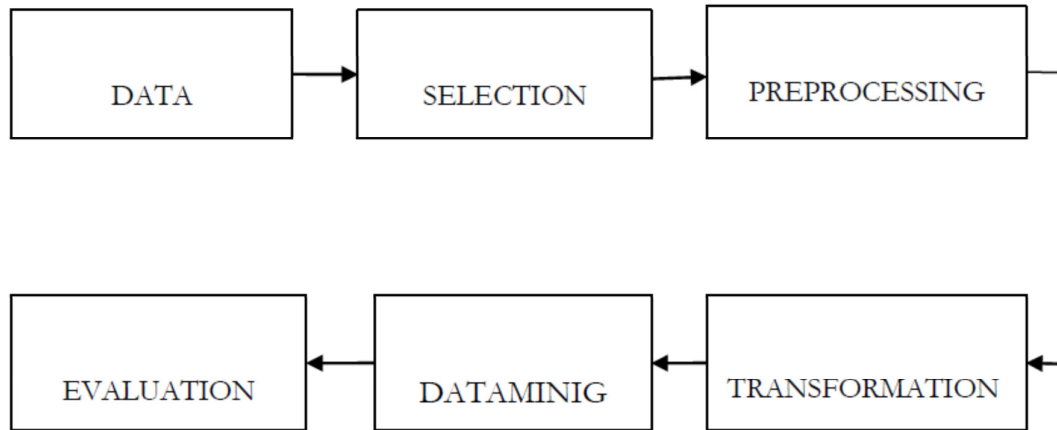
and academic data will lead to much determining factors in order to predict the student performance.

**Keyword: Data Sparsity, Attribute, Data mining Technique, Weka**

## 1. Introduction

As we enter the twenty-first century, it may not be a new message that the importance of engineering education is growing across the whole world. Technological advancements are growing. Newer and newer inventions are happening day by day.

Nowadays, data mining is playing a vital role in educational institutions and one of the most important areas of research with the objective of finding meaningful information from the data stored in huge dataset. Academic data mining (ADM) is a very important research area which helpful to predict useful information from educational database to improve educational performance, better understanding and to have better assessment of the students learning process. Data Mining or knowledge discovery has become the area of growing significance because it helps in analyzing data from different perspectives and summarizing it into useful information



**Figure 1: Step of Data mining technique & extracting knowledge**

The ability to predict a student's performance is very important in educational environments. Students' academic performance is based upon diverse factors like personal, social, and psychological and other environmental variables. A promising apparatus to achieve this goal is the utilization of Data Mining. Information mining systems are utilized to work on enormous measure of information to find shrouded examples and connections supportive in basic leadership..

This study investigates and compares the educational domain of data mining from data that come from students personal, social, psychological and other environmental variables. The extent of this exploration paper, makes to extricate the information find from the understudy database for improving the understudy execution.

## 2. Related Work

**M. Mayilvaganan et al 2014 [1]**,the aim of this study is to compares some classification techniques used to predict the performance of student. It is helps to analyse the slow learner in the semester exams that are likely study in poor which are used to improve their skill as early to achieve the goal in end semester. The task can be processed based on the several attributes to predict the performance of

the student activity respectively. In this research, the paper has been focused the improvement of Prediction/ classification techniques which are used to analyse the skill expertise based on their academic performance by the scope of knowledge. Also the paper shows the comparative performance of C4.5 algorithm, AODE, Naïve Bayesian classifier algorithm, Multi Label K-Nearest Neighbour algorithm to find the well suited accuracy of classification algorithm and decision tree algorithm to analysis the performance of the students which can be experimented in Weka tool.

**Vasile Paul Brefelean et al 2007 [2]**, Decision trees classifiers are simple and prompt data classifiers as supervised learning means with the potential of generating comprehensible output, usually used in data mining to study the data and generate the tree and its rules that will be used to formulate predictions. One of the major challenges for knowledge discovery and data mining systems stands in developing their data analysis capability to discover out of the ordinary models in data. The excellence of a university is specified among other concerns by its adapting competence to the constant changing needs of the socio-economic background, the quality of the managerial system based on a high level of professionalism and on

applying the latest technologies. This article represents an implementation of a J48 algorithm analysis tool on data collected from surveys on different specialization students of my faculty, with the purpose of differentiating and predicting their choice in continuing their education with post university studies (master degree, Ph.D. studies) through decision trees.

**Samrat Singh, Dr.Vikesh Kumar et al 2013[3]**, Data Mining is a powerful tool for academic performance. Educational Data Mining is concerned with developing new methods to discover knowledge from educational database and can be used for decision making in educational system.

**NeelamNaik&SeemaPurohit et al 2012 [4]**, The quality higher education is required for growth and development of country. Professional education is one of the pillars of higher education. Data mining techniques aim to discover concealed knowledge in existing educational data, predict future trends and use it for betterment of higher educational institutes as well as students.

**S. A. Kumar et al 2011 [13]**, Educational data mining is used to study the data available in the educational field and bring out the concealed knowledge from it. Classification methods like decision trees, rule mining, Bayesian network etc. can be applied on the educational data for predicting the student's behaviour, performance in examination etc. This prediction will help the tutors to identify the weak students and help them to score better marks. The C4.5 decision tree algorithm is applied on student's internal assessment data to predict their performance in the final exam. The outcome of the decision tree predicted the number of students who are likely to fail or pass. The result is given to the tutor and steps were taken to improve the performance of the

students who were predicted to fail. After the declaration of the results in the final examination the marks obtained by the students are fed into the system and the results were analysed. The comparative analysis of the results states that the prediction has helped the weaker students to improve and brought out betterment in the result. To analyse the accuracy of the algorithm, it is compared with ID3 algorithm and found to be more efficient in terms of the accurately predicting the outcome of the student and time taken to derive the tree.

### **2.1. Problem Statement**

Data mining is widely used in academic sector to find the problems arise in this sector. Student performance is of great concern in the academic institutes where several factors may affect the performance. For prediction the three required components are: Parameters which affect the student performance, Data mining methods and third one is data mining tool. These Parameters may be psychological, personal, and environmental. We conduct this study to maintain the education quality of institute by minimizing the diverse effect of these factors on student's performance. In this Paper, Prediction of student Performance is done by applying Apriori classification techniques WEKA tool. By applying data mining techniques on student data, we can obtain knowledge which describes the student performance. This knowledge will help to improve the education quality, student's performance and to decrease failure rate. All these will help to improve the quality of institute.

#### **2.1.1. OBJECTIVES**

- To formulate a refined dataset specific to problem domain for testing purpose.



- To achieve proper clustering of instances in unsupervised machine learning scenario when no labels or classes are provided to instances.
- To analyses the performance of clustering by classification.
- To find combination of demographic attributes for accurate classification of instances and for generating decision rules.

### 3. Weka file format

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software. ARFF files have two sections: Header information and Data information [15].

The Header of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types.

The @RELATION, @ATTRIBUTE and @DATA declarations are case insensitive. Lines that begin with a percentage sign are comments.

#### *The ARFF Header Section*

The ARFF Header section of the file contains the relation declaration and attribute declarations.

The @relation Declaration

The relation name is defined as the first line in the ARFF file. The format is:

@relation < relation name >

Where < relation name > is a string. The string must be quoted if the name includes spaces [16].

The @attribute Declarations

The format for the @attribute statement is:

@attribute < attribute name s><datatype>

Where the < attribute name > must start with an alphabetic character. If spaces are to be included in the name then the entire name must be quoted. The <datatype> can be any of the four types currently (version 3.2.1) supported by Weka:

Numeric

< Nominal specification >

String

< Date format >

#### *ARFF Data Section*

The ARFF Data section of the file contains the data declaration line and the actual instance lines.

The @data Declaration

The @data declaration is a single line denoting the start of the data segment in the file. The format is: @data

The instance data

Each instance is represented on a single line, with carriage returns denoting the end of the instance. Attribute values for each instance are separated by commas. They must appear in the order that they were declared in the header section (i.e. the data corresponding to the nth @attribute declaration is always the nth field of the attribute).

Missing values are represented by a single question mark, as in:

@data 4:4;?; 1:5;?; Iris setosa

Values of string and nominal attributes are case sensitive.

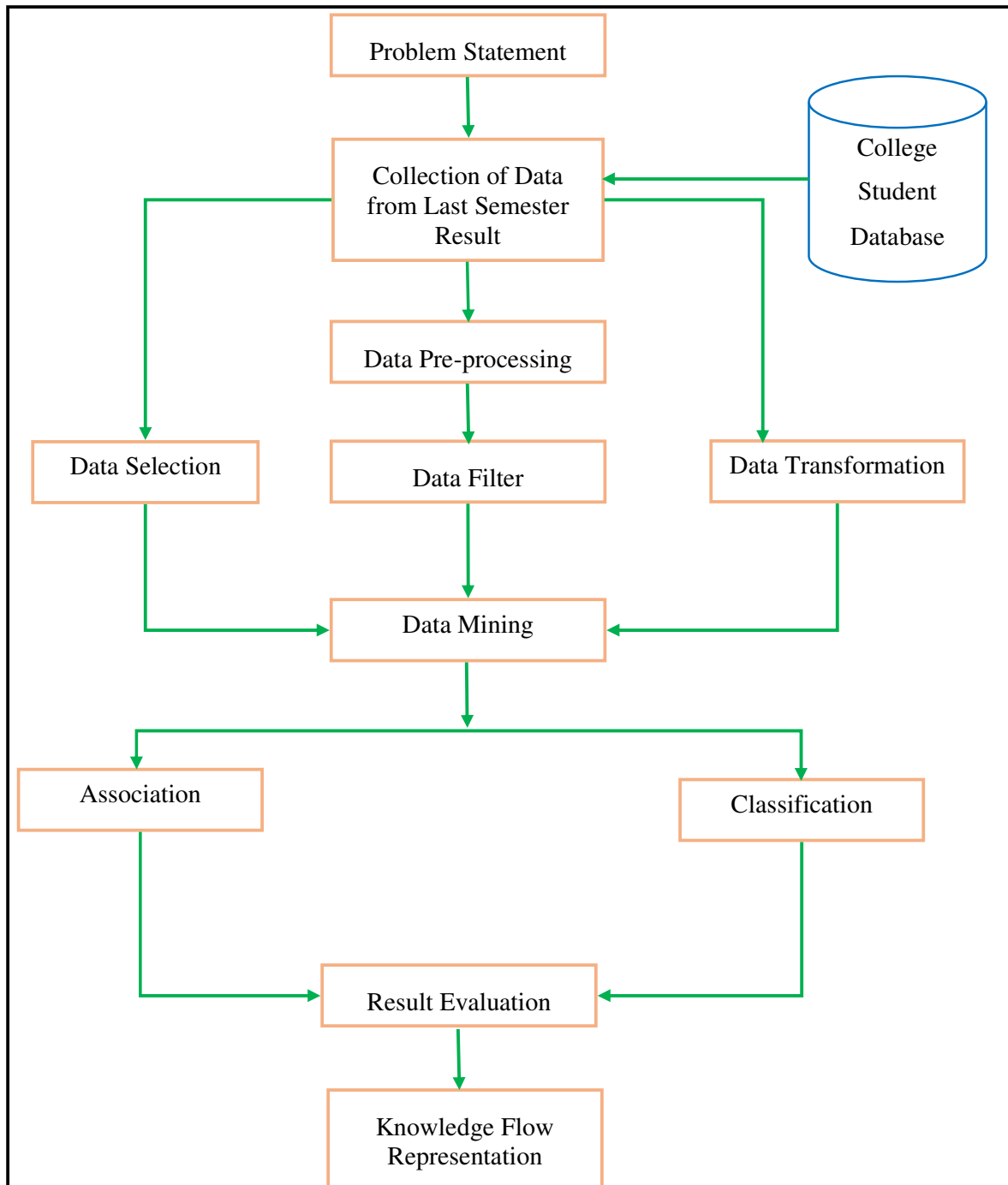
### 4. Work Flow Diagram

The main objective of this work is to use data mining methodologies to student's performance in



the semester. Data mining provides many tasks that could be used to study the student performance. Our work will be divided into two main parts- one is prediction by classification and another one is association rule mining by using the machine learning tool 'WEKA'. At first we will select our dataset and then perform pre-processing of it. After

pre-process we will do classification over the dataset and perform prediction of result. Then we will apply association rule mining technique over the dataset and generate some rules which will be analysed later. At last both result of prediction and association will be visualized by 'Knowledge Flow Representation'.



**Figure 2: Flow diagram of Data mining for Student data**

#### 4.1. Data Collection and attribute Selection

In this phase only those arenas were selected which were required for data mining. A few resulting variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response

variables which were derived from the database are given in Table I for reference.

The data-set considered consists of 318 tuples and 19 attributes [15]. Respectively tuple represents the attribute values of a student or it describes the

details of student in terms of academic performance and social behaviour.

**Table I: Attribute description and parameter Value**

Attribute	Description	Parameter Value
GENDER	Student Gender	Male, Female
CAT	Student Category	General, OBC, SC, ST
BRANCH	Student Program Name	CSE,IT, EE, EX
RollNo	student ID	Number Type
Mid Sem Attendance	Mid Sem test Attendance	Absence , Present
Mid Sem Mark	Internal marks	between 0-30
TH/Subject	Theory marks of Subject	between 0-70
PR/Subject	Practical marks of subject	between 0-50
Result	Result in branches	Outstanding >70, Excellent >60, Good>45, Pass>33, fail<33

**5. Conclusion**

This paper presents data mining in education environment that identifies students' failure patterns using association rule mining technique. The identified patterns are analyzed to offer a helpful and constructive recommendations to the academic planners in higher institutions of learning to enhance their decision making process. Association rule mining has been applied to Education systems for analysis of student result. In this research, the association rule mining technique is used to find concealed patterns and evaluate students' performance and trends. Apriori algorithm is used for finding associations among attributes.

**Reference**

[1]. M. Mayilvaganan, D. Kalpanadevi, "Comparison of Classification Techniques for

predicting the performance of Students Academic Environment", International Conference on Communication and Network Technologies (ICCNT), IEEE 2014, pp 118.

[2]. Vasile Paul Brefelean, "Analysis and Predictions on Students' Behavior Using Decision Trees in WekaEnvironment " Proceedings of the ITI 2007 29th Int. Conf. on Information Technology Interfaces, June 25-28, 2007, pp 55-56.

[3]. Samrat Singh, Dr.VikeshKumar , "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques ",IJCSSET February 2013.

[4]. NeelamNaik&SeemaPurohit, "Prediction of Final Result and Placement of Students using Classification Algorithm"International Journal of Computer Applications (0975 – 8887) Volume 56– No.12, October 2012

[5]. AlaaM.El-Halees,Mohammed M. Abu Tair, "Mining Educational Data to Improve Students'Performance: A Case Study",International Journal of Information andCommunication Technology Research, 2012.

[6]. B.K. Bharadwaj and S. Pal,"Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[7]. SuchitaBorkar, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining ", IJCSMC,Vol. 2, Issue. 7, July 2013, pg.273– 279.

[8]. Randhir Singh, M.Tiwari, NeerajVimal,"An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education", 2013.



- [9]. D.MagdaleneDelightAngelina,"Association Rule Generation for Student Performance Analysis using AprioriAlgorithm",The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 1, March-April 2013
- [10]. Mrs. M.S. Mythili, Dr.A.R.MohamedShanavas,"An Analysis of students' performance using classification algorithms ",ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver .III (Jan. 2014), PP 63-69
- [11]. S. Anupama Kumar and Dr.Vijayalakshmi M.N "Implication of classification Techniques in Predicting Student's Recital" International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.1, No.5, September 2011.
- [12]. I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.
- [13]. S. A. Kumar and M. Vijayalakshmi, "Efficiency of decision trees in predicting student's academic performance," in First International Conference on Computer Science, Engineering and Applications, CS and IT, vol. 2, pp. 335–343, 2011