STUDENT PERFORMANCE MEASURE BY USING DIFFERENT CLASSIFICATION METHODS OF DATA MINING

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Software Engineering

Submitted By

Neha Choudhary

(Roll No. 801431016)

Under the supervision of:

Dr. Ashutosh Mishra

Assistant Professor

Thapar University, Patiala

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

June 2016

Certificate

I hereby certify that the work which is being presented in the thesis entitled, "*Student performance measure by using different classification methods of data mining*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Software Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Ashutosh Mishra* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Neha

(Neha Choudhary)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Ashu

(Dr. Ashutosh Mishra)
Assistant Professor,
CSED

Countersigned by

(Dr. Deepak Garg)
Head
Computer Science and Engineering Department
Thapar University
Patiala

(Dr. S. S. Bhatia)
Dean (Academic Affairs)
Thapar University
Patiala

Abstract

The assessment in outcome based learning is very vital and significant approach toward measuring the student's performance. There are many traditional methods existing in this context. The data mining is one of the intelligent computing methods which are having widely accepted features that enable the idea of its usage in assessment. Much work has been done to measure the student performance by using different methodologies and modern technologies. In this work, we have gone through the current datasets of students of the university and different classification methods of data mining are used to measure the accuracy of student performance. Based on the analysis of the result, it has been concluded that accuracy and the other measures of SVM is more than the other classification methods.

Acknowledgement

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life. It is a great privilege to express my gratitude and admiration towards my respected supervisor **Dr. Ashutosh Mishra**, Assistant Professor, Computer Science & Engineering Department, Thapar University, Patiala. He has been an esteemed guide and great support behind achieving this task. This work would not have been possible without the encouragement and able guidance of him. I also thank my supervisor for his time, patience, discussions and valuable comments. His enthusiasm and optimism made this experience both rewarding and enjoyable. I am truly grateful to him for extending his total co-operation and understanding whenever I needed help and guidance from him. I am also heartily thankful to **Dr. Deepak Garg**, Associate Professor and Head, Computer Science & Engineering Department and **Rupali Bhardwaj**, PG coordinator, for motivation and providing uncanny guidance and support throughout the preparation of the thesis report.

I will be failing in my duty if I do not express my gratitude to **Dr. S. S. Bhatia**, Senior Professor and Dean of Academic Affairs, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable. Last but not least, I would like to thank my family for their wonderful love and encouragement, without their blessings none of this would have been possible.

Neha Choudhary

(801431016)

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

In this chapter, we have discussed about the data mining and its relevance in educational systems. This chapter includes the main motivation and objectives of the research work. The grading system and its various challenges in the educational institutes are also discussed. We have also discussed about the various modes of data and the educational environment.

1.1 Prelude

Higher education has gained importance manifolds in the past few decades. The higher educational institutes are forced to revise its scope and objects because of the private participation. The controller of regulatory body has put some guidelines with regard to infrastructure, faculty and other resources. New technologies are being developed in the field of data management and analysis due to large supply of data being present in several companies, including both private and public. The main aim of the techniques of data mining is to discover hidden and insignificant links within the information having diverse characteristics. Various techniques of data mining are being used in different fields including the educational environment. A very encouraging area to attain this objective is the usage of Data Mining (DM) [1]. In fact, classification is one of the most helpful DM work in e-learning.

Data mining has been executed well in the business applications, but its use in higher education and higher learning institutions is still relatively new. In the sector of education, educational data mining proves to be an emerging practice which is very recent and its practice is preconceived to identify and extract new and valuable knowledge from the data [2]**.**The aim is to resolve problems of research areas of education and improve the whole educational process using various statistical techniques, machine learning programming (MLP) and data mining algorithms. Educational data Mining (EDM) is a prospering practice that can be used for analytics and visualization of data, prediction of student performance, student modelling, grouping of students etc [3].

Educational Data Mining is focussed on developing methods to explore the unique and increasingly large dataset which arrives from educational sources and further employing those methods to understand the students and the environment in which they learn in a better way. Educational Data Mining (EDM) is the process to convert raw data from education systems to beneficial information which can be further be used by parents, teachers, educational developers, other educational researchers and students.

1.2 Motivation

There are numerous challenges in the higher education like increase in the number of students, global competitive education market, rising student expectations, a demand and need for new technologies, significant reductions in government funding, etc [4]. Due to these pressures and challenges, the universities are bound to re-think on how the education can be best delivered and supported. Educational data mining helps to better understand learners as well as learning, by using the information from e-learning and web-based education to explore the relevant data and using that information to better understand the students. Therefore, educational data mining assists to develop methodologies that allow to improve the overall process of education.

It has been observed that mostly data mining techniques involve large data sets to work with. But in the ambience of education, we are usually encountered with relatively small data sets containing small groups of students. Data mining techniques are also applicable to small data sets pertaining to higher education institutions. Further, the usage of these techniques in real life situations is very advantageous. In addition, it is helpful to the administrators in decision making.

1.3 Objectives

The main purpose of any institution is to improve the quality of the education and to impart managerial decision which would be beneficial to the society. Good measurement and prediction of students is one way to reach the highest quality level in higher education system. The main objective of this research work is to find the methods which can best measure the academic performance of the students in university. We have employed data mining techniques for the following purpose:

i. To predict the final grade of the students by using the obtained grades in four semesters.
ii. Improving or discovering the main attributes which can be useful for measuring the performance.
iii. Studying about the various algorithms which uses the attribute's values at its best and thus performs better in terms of accuracy.
iv. Learning the classification algorithms in detail so that the best model is discovered and attained to measure the performance of the students.

1.4 Types of Educational Environment

The environments of education are organized through lectures, class discussions, presentations, seminars, projects, training, groups, individual work etc. Such systems aim to collect information about student attendance, marks, grades, curriculum goals and individual's data [5]. With the increasing use of computers as the form of educational tool, it is quite easier and convenient for educational instructors to measure, analyse and monitor student's behavior by using their information.

1.Massive Open Online Courses

Massive Open Online Courses (MOOC) are increasing exponentially and is also in interest from the perspective of educational community [6]. MOOC is an open-access online course used at large-scale interactive participation that makes it possible for anyone having an internet connection to enroll for free. They generate bulk amount of data that further use the data mining techniques to analyze and process.

2.Learning Management Systems

Learning Management Systems (LMS) are one of the very special types of Web-based educational platform that offers a wide variety of channels and workspaces to facilitate communication and sharing of information. LMS accumulate large logging data in accordance with student's activities and generally have built-in tracking tools for students that allows the instructor to view statistical data [7].

3.Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) are those systems that provide direct customized instruction or feedback to students. An ITS generates models of student behavior and

changes its mode of interaction with each student based on its individual model [8]. ITSs record all student-teacher interaction in log files or databases.

4.Adaptive and Intelligent Hypermedia Systems

Adaptive and Intelligent Hypermedia Systems (AIHS) is one of the special kind of adaptive hypermedia in the development of educational course that is an alternative to the traditional 'just-put-it-on-the-web' approach [9]. AIHs store data regarding student models, domain models and interaction log files.

5.Test and Quiz Systems

Exams and quizzes are one of the most widely used and well-developed tools in the field of education. A test consists of a series of questions for the purpose of collecting information from receivers. The purpose of these systems is to measure the depth of knowledge of the students with respect to concepts and subjects. Test systems store information about questions, student's answers and measured scores.

1.5 Types and mode of different data

1.Relational data

Relational data is accessible by database queries written in a relational query language such as Structured Query Language (SQL). A relational database is defined as a collection of tables consisting of a set of attributes like columns or fields and generally stores a large set of tuples like records or rows. In relational databases, relational data mining technique is used. However, relational data are generally converted into transactional data before processes of data mining is applied [10].

2.Transactional data

In the context of data management, transactional data is the information recorded from transactions. Transactional data can be financial, statistical, logistical or work-related data. In fact, many well-known and traditional data mining techniques are used with this type of data, such as classification, regression, time-series analysis, clustering and association rule mining.

3.Temporal, sequence and time series data

A temporal database caches relational data having time-related attributes involving several time-stamps. A sequence database stores sequences of ordered events, with or without a concrete notion of time. A time-series database is used to store sequences of

values or events obtained over repeated measurements of time (e.g., hourly, daily, weekly, monthly, annually).

4.Text data

Text databases consist of large repositories of documents collected from various sources. The sources could be news, messages, any article, web pages, research papers, journals, books, digital libraries, e-mail, chat and forum messages. Text databases may be unstructured, such as Hypertext Markup Language (HTML) or structured or semi-structured, such as e-mail messages and eXtensible Markup Language (XML).

5.Multimedia data

Multimedia databases stores image, audio and video data. Multimedia databases must support large objects as data objects such as video may require gigabytes of storage requiring specialized storage and search techniques. Data mining has the sub-field of multimedia data mining that deals with an extraction of absolute knowledge, multimedia data relationships and other patterns implicitly stored in multimedia databases [11].

6.World wide web data

World Wide Web (WWW) provides three types of source data [12]:

- Web pages content
- Intra-page structure
- User usage data

Table 1.1 Different types of data and DM techniques

Types of data	DM technique
Relational data	Relational data mining
Transactional data	Classification, clustering, association rule mining etc.
Temporal, sequence and time series data	Sequential data mining
Text data	Text mining
Multimedia data	Multimedia data mining
World Wide Web data	Web content/structure/usage mining

1.6 Grading system in education

The process of applying standardized measurements of deviating levels of achievement in a course is known as grading system in education. The new institutions or the schools founded did not grade the performance of their students. Monastic and cathedral-based schools came into appearance into 6th century Europe. In the high medieval period, these schools got the status of university. The first university was founded in the University of Bologna.

Three centuries ago, when European universities started fostering competitions among students for prizes and rank order, only then the grades were instituted. In the early 1700s, the Cambridge Mathematical Tripos examination became fiercer than any other competition. In the 18th century, the competitions were increasingly formalized. It took the form of any tournament in which the students moved up the rungs and the ranking system by the correct answers to the questions. As this grews up, the students faced more able opponents and more challenging questions.

1.7 Criteria and challenges for evaluating grading system

Our current practices have many adverse consequences which can be sorted to evaluate the criteria for the grading system.

1.Reflect student learning outcomes

Grades do not reflect the actual learning outcomes that the students have and have not achieved. The actual meaning of the grades A, B, C or D is not known. As a result, the capability of the students is not known in real and they probably assume that they have mastered the outcomes well enough. In addition, the word "outcome" sounds like something that has already happened.

2.Uphold high academic standards

Giving too many low grades causes administrative censure. In order to avoid such thing and enhance student satisfaction, our current system incentivizes faculty to sacrifice rigor. Due to this, students fail to refine their skills, master much content, adopt lifelong learning skills and maximize their cognitive development. Not only the problem is caused to the students, but the employees also suffer. They hire and manage graduates who are not well prepared for collaboration, problem solving, communication, analytical thinking and collaboration.

3.Motivate students to learn

Our current system encourages those who have achieved high grades, despite of putting it in the least amount of time and effort. Students believe that the desirable grades are the key to a better occupational and economic future. Performance is considered to be the main thing; everything. Learning doesn't play much role in this.But ideally, a grading system should emphasize learning over grades and encourage a learning over a performance orientation.

4.Discourage cheating

In the current grading system, some students win the higher education game by using the popular strategies of cheating and plagiarism. They have very little incentive to learn or do high-quality work. In fact, some students put plentiful time, effort and money to implement these strategies.

5.Motivate students to excel

Students are discouraged from aspiring to achieve excellence as the credits are allocated partially even when the work is less satisfactory. In fact, even the minimal effort and time is not invested. It lets the students not to focus on their course work. Our grading system should build incentives for the students to aim high, work hard and do their best. The inferior work should be marked unacceptable and strong performance should be demanded to have any points.

6.Reduce student stress

In higher education, students stress of having low grade because low grades may affect their precarious future job prospects, status and financial well-being. In this way, our system apparently proves to be hazardous to student's mental health as they feel that they lack control and grip over their academic success. This increases their anxiety because they fail to understand instructors' expectations. It is true that event the ideal grading system fails to eliminate the student stress, but a better grading education system would give more clear picture of the expectations of the faculty and parents, strong hold on their educational lives, a more control on what they can and cannot satisfactorily, better choice and volition. All such practices would be able to minimize the conflicts between the students and the instructors related to their grade.

7.Save faculty time

Grading involves ample amount of time for the faculty. The time is unpleasantly spent on passing judgments, removing partial credits, justifying the decisions to the students. The results of the grading often creates a wedge between the teachers and their students. The instructors thus spend a lot of time to write the explanations and justifications on the work of their students.

8.Make students feel responsible for their grades

Most of the time the students blame their lower than expected grades on an external locus of control. They believe that the grades are low due to the fate or their teachers. Therefore, we need a better grading system which would provide the students with clear choices and the extent up to which they should master the course material and bound their choices respective to the significance of the grades. In this way the students would realize that it is their responsibility that what they have to do for each grade.

9.Minimize conflicts between faculty and students

The stress caused to students due to grades causes the protest among the students and faculty and thus pressuring faculty to give them more points, even when not justified. This grade-grubbing not only robs the productive time of students and faculty but also harms the rapport and trust between both of them. It is possible that the students would stop confronting their instructors with grade protests if they understand the work required for each assignment and test. In a better grading system, there should be a clear and defined role in determining the grades so that the students evaluate their faculty more gently.

10.Give students feedback they will use

In the present scenario, students hardly view the constructive advice on how to improve next time. They seem to take the feedback of the faculty as justification for the grade. The feedback should aim to maintain the quality of the education and be considered as constructive and useful.

11.Grading rules be simple

A grading system should simplify the point system and faculty decision making. Sometimes even the instructors get confused with too many rating levels for too

many individual assignments and tests. The students often forget the details about how they will be graded in a course even after asking so many questions about the grades.

12.Have high inter-rater agreement

With the increase in the number of rating gradations, the disagreement among raters also increases. This is true when the details about the expectations for the work are not clear and explained in detail. One solution of this is to use a rubric in a uniform way. Another one is to use a better grading system. The standards of the assessment should be clear enough so that the colleagues agree on their rating of given student products.

1.8 Ideal grading system

We employ the constructive use of grades once they are made more than information. The grades that are made merely to sort humans any partial way is generally abused and is unhelpful in the long run. The main elements for building successful grading systems are as follows:

1.Teacher utility

Teachers are the main stakeholders for using the grading systems. The grading system is not considered to be wonderful if the teachers can’t use it effectively. It need to be manageable, especially for those who use it daily.

2.Transparency

The performance levels are clearly defined at every level. It must be certainly clear at every entrance into the system. Some comments and pop-up textboxes should be revealed listing the criteria for evaluation about the student's performance.

3.Evidence-based

Grades must be used as the way to report the students' performance against standards, nothing else. The teachers should be properly trained to calibrate evidence of standards. Good grading systems have probity because they are based on evidence of standards. They are not dependant on the routes students walked in to get there, the inconsistent emotional state of their instructors or the non-uniform classroom management techniques.

4.Feedback focussed

It is possible for students to learn without grades but the learning becomes more effective if timely descriptive feedback is provided. The teachers are not properly trained to provide proper feedback. Detailed feedback experience is an overt act of direct learning. But many teachers fail to realize this fact and think that they need to stop teaching in order to provide proper feedback.

5.Disaggregated

It is more useful to have fewer curriculums aggregated into one grade or symbol. The standards being assessed are recorded with their separate scores. The larger topics are broken down into further subcategories.

6.Mode of evaluation

Student's final grade does not determine the standards-based grading. The grading system is effective if it do not average scores. The averaging may distort the accuracy of the final grade. Effective grading systems require that the performance regarding important standards must be checked repeatedly. The teachers should put previous curriculum on subsequent tests.

7.Constructive response to failure

There is a need for constructive responses to late work and re-doing assessments and assignments. The teachers should be highly trained to study how the mind develops. In case the student does not perform well, the faculty should response with concern and more effort towards his study.

1.9 Thesis Organisation

Our research work proposes an effective methodology for measuring the student's overall performance by using the grades of each semester. Different results of the classification algorithms of data mining are analyzed and final outcome is made based on the accuracy of the model. The rest of the thesis is as organised below:

The **second chapter** gives an account of the literature survey and the various approaches of data mining in higher education systems. The **third chapter** deals with the problem statement. This chapter deals with the main aim of carrying this research work and the objectives of the thesis. The **fourth chapter** identifies the tools and techniques employed to do the proposed work. This chapter acquaints with

the IBM SPSS Modeler and its features, the data mining tasks. The predictive data mining model: classification, regression and time series analysis and the descriptive data mining model: clustering, summarization and association are also explained. In **fifth chapter**, we present the methodology on the proposed approach. This chapter includes the description of workflow, the three algorithms used: Support Vector Machine (SVM), C5.0 and Bayesian network. The four kernels of SVM are also explained in detail. In the **sixth chapter**, we have shown the implementation and results to measure the student's performance. It includes the procedure applied, experimental setup, results and observations. The **seventh chapter** concludes the present work done under different constraints and provides possible improvements in future.

Chapter 2
Literature Review

In this chapter, we discussed about various approaches for student performance prediction and measurement using various techniques of data mining, exercised by various researchers.

From several years, various data mining classification and clustering models have been constructed and executed to analyze and measure the performance of students. For instance, AHP has been employed successfully to predict the student course selections in higher educational institutions and the outcomes reveals that the accuracy of the student's course prediction is quite good [13]. Course selection is a very vital decision because on this very decision, the life of the student is dependent. Many models are generated with regard to this thing but very few models examined that specific academic term and academic year are also very important in which the students will opt those courses in the future.

Therefore, it is necessary that student's preferences and interests should be present in the student information systems of the institutions. The dataset for the study was collected from Canadian research-intensive university for students pursuing an undergraduate degree program. If the data is fully identifiable, then the accuracy for predicting the higher course selection is very high.

2.1 Approaches of Data Mining in Higher Education

Shahiri et al. [14] have also provided an overview on several techniques of data mining that were applied to predict and analyze performance of students, concentrating on the identification of most valuable attributes in a student's data by employing the prediction algorithm**.** They provide a systematic literature review to improve the student's achievements by using the techniques of data mining. The various analytical methods used cumulative grade point average (CGPA) as their data sets, thus helping the system of education to monitor the performance in a very systematic way.

Osmanbegović and Suljić [15] applied three supervised data mining algorithms to assess the data of first year students to predict favourable outcome in a course and

evaluating the performance based on certain factors like convenience, accuracy and approach of learning**.** A very high emphasis is given on some socio-demographic factors, high school results obtained, attitude towards study and marks in entrance examinations. The whole data was collected from University of Tuzla, academic year 2010-2011. The authors believe that exams play a very important role to determine the future of the students, in addition to the internal assessments. They used WEKA for their study and implemented it in java and also conducted four tests to assess the input variables: Info Gain test, Chi-square test, Gain Ratio test and One R-test.

A model has also been developed based on some selected input attributes assembled through questionnaire method [16]. Ramesh et al. conducted a survey cum experimental methodology to generate database for the students for predicting the performance. The three main objectives were to identify the essential predictive variables on higher secondary students, know the best classification algorithm and to predict the grade at higher examinations. The study shows that parent's occupation plays a major role and not the type of school in predicting the grades. The data for the study was collected from schools and internet and the authors found out that multilayer perceptron algorithm is the best one for grade prediction. This algorithm is more efficient showing the accuracy of 72%.

Goga et al. [17] designed a tool by using .NET framework to predict student's grade by providing various parameters as input. Models based on the student's enrollment records were developed by using ten classifications trees (OneR, Random forest, ZeroR, random tree, Decision stump, REPTree, JRip, , J48, PART, and Decision table) and a multilayer perceptron (Artificial Neural Network) learning algorithms by operating on WEKA(Waikato Environment for Knowledge Analysis). A framework is designed for intelligent recommender system which recommends suitable actions for improvement. The work is based on the background factors that predicts tertiary first year academic performance of the students. The data for the study is taken from Babcock University, Nigeria. The background factors for the students were collected through in-depth interview. The various demographic factors are father occupation, mother occupation, family income, place of birth, family size, academic qualification of parents, parent's marital status. The benchmarks used in the comparison of the generated models include confusion matrix, accuracy and speed. Random tree outperformed the other algorithms in terms of benchmarks. Therefore, random tree is

adopted as the best algorithm in the domain of this study to serve as a building block for designing a generic system.

Prediction model is also developed based on the participation of the students through Genetic Programming by integrating learning analytics and educational data mining [18]. The author integrates these to build the final performance prediction model by the usage of interpretable Genetic Programming. The data is collected from an environment of a collaborative geometry problem solving known as Virtual Math Teams with Geogebra (VMTwG). Genetic programming approach is used which outperforms the traditional methods in prediction rate and interpretability. This prediction model is selected because it represents the most optimal trade-off between model understand ability and the predication accuracy. This study provides us the prediction model which is quite practical and interpretable. The four approaches integrated for modeling are: EDM, learning analytics, applied practices and HCI theory. The results reveals that the model based on genetic programming is highly interpretable and offers much optimized prediction rate as compared to traditional modeling algorithms. Practical recommendations are also outlined which selects the best prediction model among the other available algorithms.

Although many studies are being carried out on the prediction of student performance, but very few studies focus on investigating how the performance of students evolves during their course of study. The recent research by Natek and Zwilling[19] compares two tools of data mining applied to data sets of small size related to institutes of higher education and summarizes that the results will encourage the institutions to incorporate the methods of data mining to be an essential segment of higher education institutes and intelligence management systems. The author successfully predicts the success rate of the students enrolled in the trade. According to them, data mining is better approach than OLAP and statistical tools to identify the hidden patterns and data prediction.

R.Campagni et al. [20] presents the methodology to determine the future career of university graduate students. The main aim is to identify the strategy to improve the performance and scheduling of the exams by using various approaches of data mining. “Ideal career” is introduced which is basically the career of an ideal students who has given the exam just after the end of studying the particular course i.e. giving

exams without any delay. The methodology is applied to a real case study and it has been observed that the performance, in terms of final grade and graduation time, increases manifolds if the student follows the order as given by the ideal career. Bubblesort distance is used to calculate the career distance between the normal student and the student following the ideal career. The practical implementation of the obtained results has been used for the students enrolled in the Computer Science and engineering department at the University of Florence, helping them to improve.

C.Romero et al. [21] compares various methods of data mining tools and techniques for classifying students on the basis of their Moodle usage data and the final marks secured in their corresponding courses. Real and factual data is fetched from seven Moodle courses from students of Cordoba University. The authors states that a classifier model should be both, comprehensible and accurate for instructors for decision making which is actually appropriate for educational purpose. Also, discretization and rebalance pre-processing techniques are applied on the numerical data to justify whether better classifier models are attained. The training knowledge acquired, either for the specification of the teaching scenario, or for the creation of the student model, finding the deviations of student's behaviour is one of the unresolved obstacle faced in the development of intelligent tutoring systems [22]. Sequential pattern mining and constraint relaxations can be employed to automatically get the knowledge. Manolis Mavrikis[23] shows the enhancement of two machine-learning models which are used to predict whether a student can answer right questions in ILE (Interactive Learning Environment) without requiring any aid and whether a student's interaction is useful or not in terms of learning. The author uses ICS (Improved Cuckoo Search) algorithm of WEKA to obtain Bayesian network which provide appropriate prediction. A computer agent named "Betty" using a visual concept map portrayal is used to build models and analyze behaviour of students in different versions of learning [24].

Kaur and Kaur [25] worked on the new system of multistage examination in higher education institutes introduced by UGC (University Grants Commission), India. The authors used data mining techniques to predict and examine the performance of 1000 students. Their study proved to be helpful to the community for understanding the behaviour and learning of students in terms of the difficulty levels. According to them, Classification and Regression Tree (CART) is observed to be the best

classifiers to predict student's grades, supplemented by AdaBoost. As the difficulty level of the subjects varies, the accuracy of the prediction also varies. There are several problems in the higher educational institutions that are very hard to solve manually so such sort of study helps the management, the parents and the teachers to study the behaviour of students and take useful decisions.

R. Asif et al.[26] explores how the performance of the students evolves during their years of studies. For clustering, progression of students is used i.e. the students in same cluster have same progression. For the study, the data is taken from NED University of Engineering & Technology (NEDUET), Karachi for IT bachelor degree students. Two consecutive cohorts have been analyzed by using X-means clustering. Rapid miner tool is used to cluster using X-Means and K-means algorithms. When the K-means algorithm is modified to automate the estimation of number of clusters in an optimal way, it becomes X-means algorithm [27]. Two consecutive cohorts have been examined by using X-means clustering algorithm.

Harwati et al. [28] maps the students by employing K-means clustering algorithm. This mapping should be made before the performance improvement program is designed. Therefore, this algorithm reveals the hidden patterns and successfully classifies the students. Many demographic factors such as sex, origin, CGPA, percentage etc, are taken into account while covering the data of 300 students. The authors used SPSS16 for their research. There are two types of clusters: student's academic profile as GPA, lab grades, lecture grades etc and student's activity profile as participation, origin area, attendance etc.

Mativo and Huang [29] adapted the methodology to be used for small dataset, 48 students enrolled in engineering. They developed and tested two algorithms: Support Vector Machine and Multiple Linear Regression. The three criteria for evaluation include accurately predicted grade range, mal-alarm and missing alarm. The results examined showed that SVM models produce higher accuracy to identify the students having low grades. It is used to predict the range of student's grades. More often, the MLR model over estimated the performance. It failed in number of cases and could produce correct results in only quarter of cases.

Mishra et al. [30] uses those factors to build performance prediction model which has not been used by anyone so far. Various classification techniques are used on

students' social integration, academic skills and emotional skills. The data is collected from Guru Gobind Singh Indraprastha University(GGSIPU) for MCA students, to predict their performance in 3rd semester. Two algorithms: J48 and Random Tree have been employed for early prediction. The reason to consider third semester is that most of the students are observed to drop out of course after their first year. In addition, the students normally take a year to integrate in the environment of any academic institute. The authors aim to study the impact of various factors in predicting the performance of the students. It has been found out that the result of second semester strongly influences the result in third semester, especially the programming subjects. Also, the consistent good performer performs well in the third semester too. Leadership quality also has a very valuable impact on the performance of the students. The results reveal that random tree has higher accuracy than J48 algorithm.

Garcia and Mora [31] believes that it is important to know the level of the student preparedness before at the time of taking the admission. The main purpose is to obtain a model for the first year students to predict academic performance while taking into account different academic and socio-demographic variables. The data is fetched of the first year students studying in School of Engineering. Data mining techniques are used to obtain about 60% accuracy using Rapid miner software. The data was collected from the survey related to their socio-demographic variables. The prediction increases when naïve Bayes classifier is used. The results are helpful for taking appropriate decisions even before the courses commence.

Artificial neural network also serves as a vital means to predict the performance of the students. Arsad et al. [32] conducted this study on the students studying in Universiti Teknologi MARA (UiTM), Malaysia. Cumulative Grade Point Average (CGPA) is used for the electrical engineering students to measure their academic achievement in 8th semester. The study is carried on for two entry pints as the inputs: matriculation and diploma. To measure the performance, Correlation R and Mean Square Error (MSE) are being used. The final CGPA is predicted and it is observed that it is strongly influenced by the core subjects of semester first and semester third. During this exercise, the input data was split into either the training set or the testing set. The study proves that there is a strong connection between the subjects studied at the early semesters and the final grade obtained in the last semester. Therefore, it is very essential that the concepts of the fundamental subjects should be very clear and must

be fully understood, made clear and grasped because without doing that the other subjects in the next higher semester would be very difficult.

Decision trees can be employed successfully to evaluate the student's data which is of great concern to the higher education [33]. Entropy is calculated taking the attributes into account of the educational dataset. The attribute which has the highest information gain is considered to be the root node after which the further tree is split. Therefore, more emphasis can be given to the students who have performed poor in the class. To analyze and predict the result of the class, two important factors are sessional marks and class performance. The trees are generated after studying the training set thoroughly and are used to formulate the predictions [34]. In the whole process, WEKA is used as simulation tool and the results are evaluated by using decision tree classifiers. The result reveals that poor performance in the sessionals comes out to be the major factor of student's failure in the final exams.

Bunkar et al. [35] applied the data mining classification techniques to improve the quality of educational system by examining the student's data. Authors have collected the data from university and analysed the main attributes which affects the performance of students. An automated system is developed which uses decision tree and generated rules to predict grades of students. They examined some decision tree algorithms such as ID3 (Iterative Dichotomiser 3), C4.5, CART (Classification and Regression Trees). The collected data is pre-processed to find out the parameter and variables which affects the study. Their model predicts the students most likely to fail and the students to be considered for counselling.

According to Parmar et al [36] higher educational students performance plays a vital role in career. In today's digital world a large amount of educational data is stored, which can be used to analyze the performance. Authors have used the classification methods to predict performance of students in distributed environment. The results of the model which are at local level, is difficult to extract the knowledge at global level. So to make decision making more effective it is crucial to generalize the information of the models, some of the classifier can be used for generalization. They have applied Random tree algorithm which gives higher accuracy in single data set. To obtain higher accuracy on distributed data set, rules from all data sets needs to be considered. Results shows that distributed training and testing classification on each nodes and

central node respectively, enhances the performance of prediction task on large data set. By using this prediction results and giving proper training to weak students, their performance can be improved.

The data obtained from predicting the performance of students in academic field and teachers can be used to give extra attention to students and decrease the failure rate. They use distributed data mining methods to extract the data and then apply clustering and classification techniques on it. It provides an efficient and fast solution for storing and accessing data. Clustering is applied on data to get more efficiency on process of monitoring the student performance in engineering graduation to get accurate results. After this the classification is applied to develop prediction model by using student skills and past performance in exams. By using distributed data mining, this method becomes independent of the location of data, it serves when the request comes. It has a middleware that maintain Meta Data directory, which contains server information about the location in database, and server accordingly. It also stores previous request and results, which will be used if same request comes. In case of a new request, suitable server is assigned to processing [37].

Huang and Fang [38] presented a mathematical model for making early prediction, which works even before the semester starts, and predict the score of student in final exam and engineering dynamic courses. They applied four different mathematical model multilayer perceptron neural networks, support vector machine, multivariate linear regression and multilayer perceptron neural networks. Experimental results depicts that there is no significant difference in accuracy of these model within five factor used in the study. Data of around 1900 engineering students were collected of four semesters. The model was developed using the results of first semester and validated using next three semester data. The following criteria were employed: Average Prediction Accuracy which specifies the ability of model to accurately predict exam score of students in class; Percentage of Accurate Prediction (PAP), which is the ratio of total number of accurate prediction and total predictions. Experimental results imply that any of the four applied model can be used for predicting student's score. Few of the other important factors which affect the prediction are learning style, interest, motivation, family background, teaching effectiveness and learning progression.

Hidden patterns in student's data can be extracted using data mining methods. These patterns can be utilised to determine the factors affecting the student's performance. Patil and Mane [39] proposed a General Sequential Pattern Mining Algorithm to find common pattern from student's data and used tree algorithm to construct the tree using these pattern. The constructed tree can be used to predict performance of student and if one is found on risk of failure, can be provided help accordingly. Collected data were pre-processed which include data integration, cleaning, selecting attributes and data transformation. Pre-processed is fed to GSP algorithm which finds general sequential patterns present in data set and are responsible for affecting performance. It is apriori algorithm having join and prune phase. In the final step of pattern mining, common candidate sequences are discovered. FP-Tree Algorithm has been used to generate tree of common sequences which can be used to predict high and low performance students.

Chapter 3
Problem Statement

3.1 Research Gap Analysis

Examination has an important role in the life of students. The result decides the future of the students according to the marks obtained in examination. Therefore, prediction of student's result, pass or fail, in any examination becomes very vital. More efforts can be taken to improve the studies and the performance if the student is expected to not perform well. This will help the students to pass the examination and improve performance [40]. Although many studies are being carried out on the prediction of student performance, but very few studies focus on investigating how the performance of students evolves during their course of study. Most of the approaches have used only the factors like demographic factors, academic marks as their basis of prediction. Very few work is done which take the grades of each semester into the account and work solely on the grades obtained by the students.

3.2 Problem formulation and approach

Performance of the students is calculated so that the results can be improved and the future of the student is secured. Therefore, it is very necessary that the marks or grades in each semester be taken into consideration so that the student considers each subject of equal importance.

Our research work proposes an effective methodology for measuring the student's overall performance by using the grades of each semester. Different results of the classification algorithms of data mining are analyzed and final outcome is made based on the accuracy of the model.

3.3 Objectives

i. To study existing techniques and tools for better understanding of concepts of education data mining process.
ii. To study and implement the various algorithms which one uses the attribute's values at its best and thus performs better in terms of accuracy.
iii. To evaluate student's informational data to determine the student performance in subject courses.

iv. Analyzing and predicting the academic performance of the students using their grades attained in the semesters using data mining.

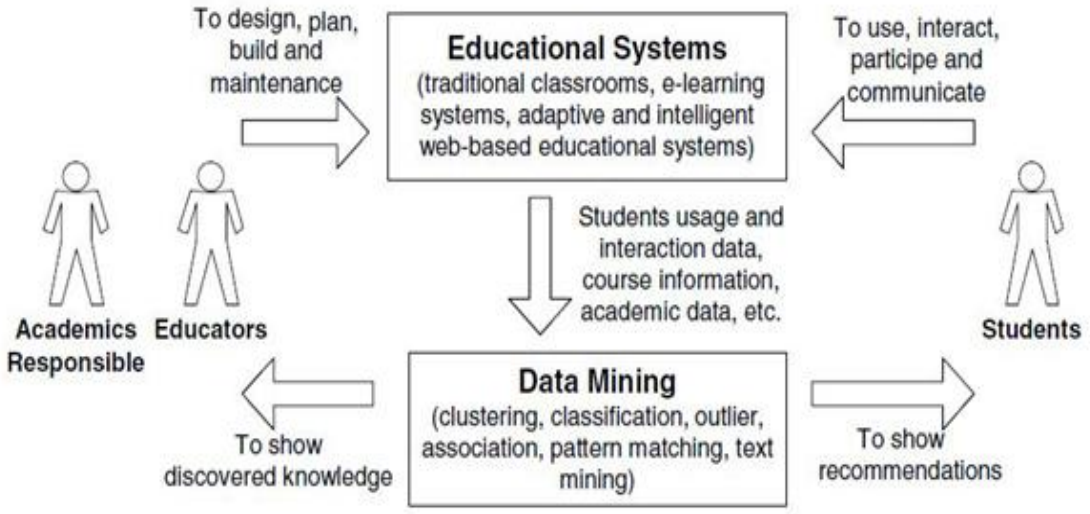


Fig. 3.1. Educational data mining cycle [1]

3.4 Scope of the research

Educational data mining has wide variety of applications in the area of student's future. It can be adopted to help the teachers as well as the students to enhance the quality of learning and student's performance by taking significance decision at right time. The various other data mining techniques and algorithms can be applied so that the results are more accurate.

Chapter 4
Tools and Techniques Used

4.1 Tool used

For the purpose of this research work, IBM SPSS Modeler is employed which is an extensive predictive analytics platform and is used for predictive intelligence decision making. It includes a wide range of advanced algorithms and techniques that helps to make the right decisions [41].

4.2 IBM SPSS Modeler Features

IBM SPSS Modeler integrates predictive analysis with optimization, business rules and decision making in the processes of the organisation helping the users and the systems to make correct decisions each time.

1.Analytical decision management

Decision management helps to integrate business rules into the processes of an organization and predictive analytics for optimization and automation of high-volume decisions at impact. The transactional decisions are automated and optimized by linking predictive analytics and business rules to output recommended actions in real time scenario.

2.Automated modelling

This method uses different modeling approaches in a one run and then compare the results of the various modeling methods. This tool helps to select which models to use in deployment, without having to run them all individually and then compare performance. Choose from automated modeling methods: Auto Classifier, Auto Numeric and Auto Cluster.

3.Text analytics

The information is usually included from unstructured text data like customer feedback, blog content, emails, web activity, social media comments.IBM Modeler is also capable for the analysis of structured numerical data and capturing the key concepts, sentiments, trends and themes. Therefore, it ultimately improves the efficiency of the predictive models.

4.Entity analytics

This feature helps to improve the coherence and flexibility of data. The entities are resolved even when any key value is not shared among the entities. The fields like customer relationship management, national security and fraud detection have a very vital importance of identity resolution.

5.Social network analysis

The relationship between social objects and the associations of these connections on an individual's behaviour is examined by social network analysis. The ones who are concerned about attrition (or churn) in telecommunications and other industries find it very useful. Predictive models are improved with the social behavior data by recognizing groups, group leaders and if others will be affected by their influence.

6.Geospatial analytics

The analytics that explores the connection among data elements which are bounded to any geographical location is known as geospatial analytics. Predictive accuracy is improved as it combines the current and historical data. Such information reveal deeper awareness about people and events. Applications such as disease surveillance, law enforcement and building and facilities management frequently uses the concept of geospatial analytics.

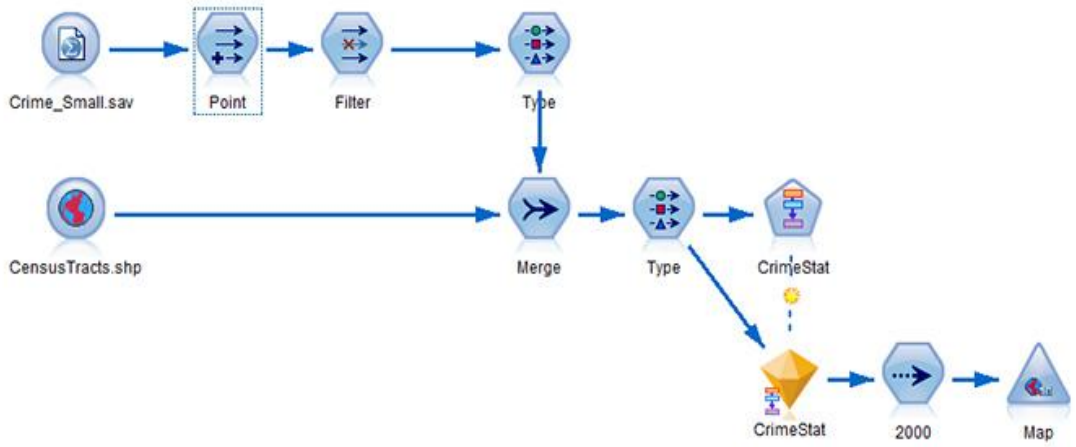


Fig. 4.1. Geographical locations using modeler

7.Modelling algorithms

The various modeling algorithms present in SPSS Modeler are as follows:

A-priori, Bayesian network, anomaly detection, CHAID,C5.0, CARMA, K-means, support vector machine, regression model, KNN, sequence, decision list, factor/PCA, generalized spatial association rule, neural networks, self-learning response model(SLRM), spatial-temporal prediction(STP), time series, two-step clustering.

8.Deployment

It is important to provide outcome to people and processes on a schedule and in real time. It helps organizations to recognize the complete advantage of predictive analytics. Deployment branches the gap between analytics and action by supporting better decision making in various areas.

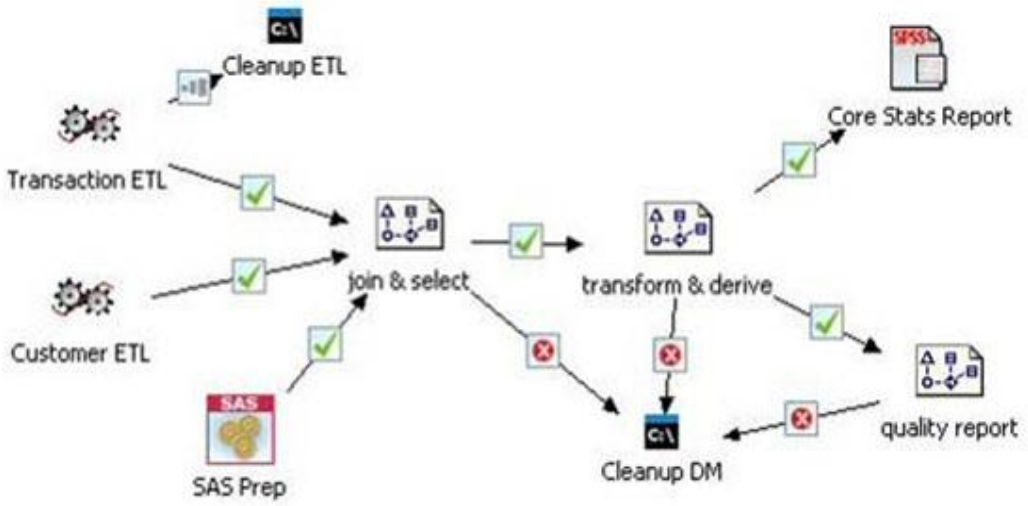


Fig. 4.2. Various deployment scenarios and reports

4.3 Data mining tasks

Data mining is a computational task for obtaining useful knowledge from the data and process that data which is successfully applied in many areas [42]. The techniques of data mining are used for building models according to which the not known data will help to establish new information. It can be predictive or descriptive by nature [43].

The Predictive model is the process of having a prediction about details of data by using known results gathered from various datasets. The Predictive data mining model have regression, classification, analysis of time series and prediction. The descriptive data mining model pin-points relationships or patterns in data sets. It proves to be fair way to investigate the characteristics of the data that is already examined and not to predict new characteristics. The Descriptive model includes task to perform clustering, sequence analysis, summarizations and association rules. The classification

techniques of data mining task are Predictive and Descriptive with their own methods is shown in below Fig. 4.3.

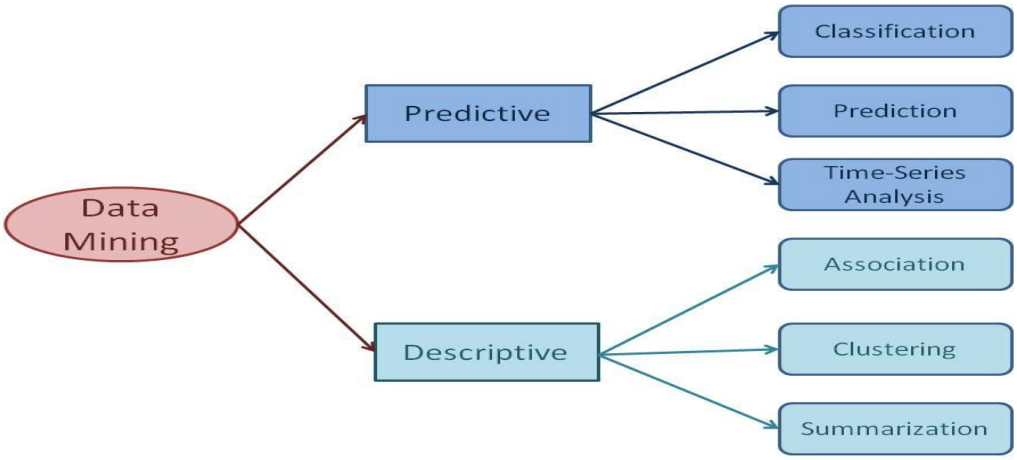


Fig. 4.3. Data mining techniques[43]

One common feature shown by all the techniques, regardless of the origin: dependencies among the attributes in the observed data and discovery of new relationships automatically. In classifying the data according to class, the algorithms are divided into two groups:

- unsupervised algorithms
- supervised algorithms

In **unsupervised data mining algorithm**:

- no supervision involved
- mining undirected
- goal is to disclose data patterns in the input fields
- main task is discovering automatically hereditary patterns in the data
- no former information about which class the data belongs to known[44]
- Eg: clustering and association rules

In **supervised data mining algorithm**:

- the class in which the data belongs for building models is familiar
- it is the method of acquiring a function that maps the data into predefined classes
- there is a particular pre specified target variable
- training data includes both the input and the desired results.

- Eg: classification

4.4 Predictive data mining model

The intention of predictive model is constructing models using support vectors, rule set, neural network and decision tree to predict the class of a recent data set as future outcome. These models study recent and historical data, thereby allowing miners to make predictions about the future. All predictive data mining models are probabilistic in nature and can only forecast what might happen in the future. When we need to prescribe an action with this model, the business decision-maker may take this information and act further. As a result, based on the enrollment information, the predictive data mining will be able to predict cluster where the future student shall fall. The various predictive data mining approaches are as described below:

4.4.1 Classification

A data mining approach that allows items in a collection to target categories or classes is known as classification. The chief goal of classification is to predict the target class for each case in the data accurately. Eg: a classification model could be useful for measuring the performance of the students to be high, medium or low. It is considered to be the "best-understood" technique among all data mining approaches. A classification task initiates building data for which the target values or class assignments are known.

Table 4.1. Classification algorithm comparison

Feature	Naive Bayes	Adaptive Bayes Network	Support Vector Machine	Decision tree
Speed	Very fast	Fast	Fast with active learning	Fast
Accuracy	Good in many domains	Good in many domains	Significant	Good in many domains
Transparency	No rules (black box)	Rules for Single Feature Build only	No rules (black box)	Rules
Missing value interpretation	Missing value	Missing value	Sparse data	Missing value

4.4.2 Regression

Statistical Regression is one of the predictive data mining model that analyzes the dependency within attribute values, that are dependent on the values of other attributes. This is the main difference between regression and classification. In other way, target attribute containing continuous (or floating-point) values requires a regression technique. The most commonly used regression type is linear regression. In this, the line that minimizes the average distance among all the points from the line i.e. that best fits the data is calculated.

4.4.3Time-Series analysis

Time-series database includes sequences of values or events acquired within repeated measurements of time. It is a sequence database having values that are typically measure at equal time interval such as weekly, hourly, daily. Time-series analysis is the process of analyzing time series data for extracting meaningful statistics and other characteristics of the data. This time series obtains the represents sequential measurements by collecting the values. It can be helpful for observing natural phenomena like wind, earthquake, treatments, atmosphere and temperature.

4.5 Descriptive data mining model

This approach for mining employs techniques of clustering, association rules mining etc. to find patterns that are covered in large data set and further aid in intelligent decision-making. As the name implies, these models "describe" or summarize raw data. They are the models that describe the past and are interpretable by humans. Descriptive data mining are useful because they allow learning from past behaviors and understand how they might influence future outcomes. It can also be used to find relevant subgroups in the bulky data. The various descriptive data mining approaches are as described below:

4.5.1 Clustering

Clustering is one of the most important descriptive data mining models. Clustering is the process of discovering natural groups (or clusters) in a database. The data items in the set in a cluster have similar characteristics. The goal of clustering is to find clusters of high quality so that the intra-cluster similarity is high and inter-cluster similarity is low. Clustering models divide the data into groups that were not defined

before. Clustering is useful for exploring data, for anomaly detection, to find natural groupings and do not use a target.

4.5.2 Summarization

The other names of summarization are abstraction or generalization of data. The aim of this technique is to map the data into subsets having simple descriptions. The data after summarization gives overview of the data with aggregated information. As the name suggests, it is that concept of data mining which involves the concept of finding a compact description of dataset. Summarization can be viewed from different angles and can be scaled up to different levels of abstraction. The main approaches of summarization are standard deviation, variance, mean, median, tabulation and mode. The applications are usually involved in data visualization, data analysis and automated report generation.

4.5.3 Association

To find relationships between attributes and items, associations or link analysis technique is used. Association rule is one of the important techniques for market basket analysis. This is so because all possible combinations of product groupings can be explored [45]. Therefore, it can be easily used to establish statistical relationships among various interdependent variables of a model. By using if/then statements, association rules helps to uncover relationships among unrelated data in various databases such as transaction oriented database, information repository or other relational database. Association rules can be used to analyze and predict the behaviour of consumers, catalog design, basket data analysis, store layout and product clustering.

Chapter 5
Methodology

An academic curriculum in the institution generally defines a specialized learning plan and schedule which put some sort of restrictions on how the students should study the courses. Many universities adopt the grading system to estimate and decide the performance of students in academics. The approach adopted by us also uses the grades for the analysis and measurement of performance.

5.1 Proposed approach

The first and foremost step is to collect the dataset required for the study. The methodology is applied to a factual data having information about the students who did their graduation in Computer Science and Engineering at Thapar University, Patiala, Punjab (India). The workflow of the research is shown in Fig.5.1.

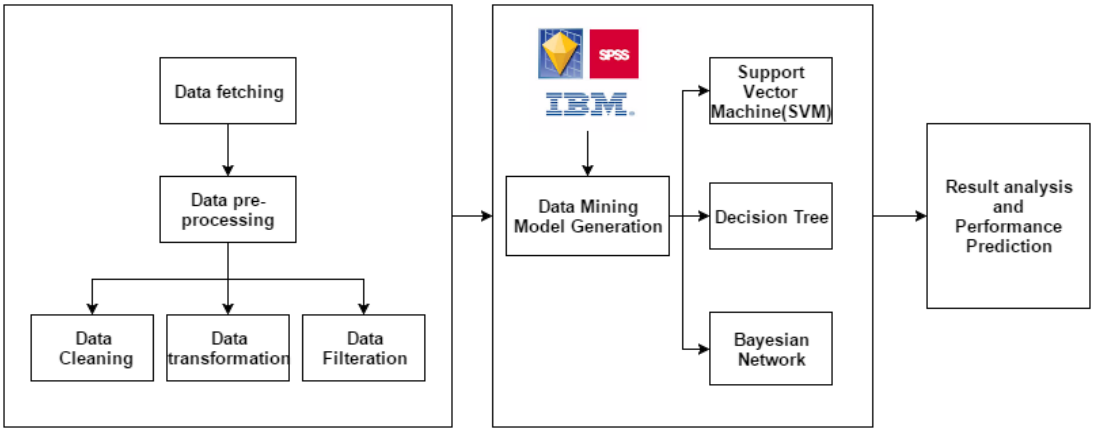


Fig. 5.1. Workflow of study

Once the data is fetched, it is transform into required form for the mining process, which is known as pre - processing phase. It is a crucial step used in data mining system that aspires to transform the raw data into a proper format for resolving a particular problem. This task is accomplished by using a certain mining method, algorithm or technique. It has been observed that finer the pre-processing is done of the raw data, the more useful and suitable information is possible to discover. This phase usually consumes 60 to 90 % of the time, training, efforts and resources employed in the complete knowledge discovery process. Asif et al. [26] highlights the importance of this step.

After the data is pre-processed, we then identify the incomplete, incorrect, irrelevant and inadequate data from our dataset and remove this erroneous and improperly formatted data. This phase is known as data cleaning phase. This process usually includes removing the typing errors or validating and correcting the values of entities or attributes by cross checking it with accurate data set. There are many attributes which have no role in our study and therefore removed like tagging for, semester type, program code, section branch etc.

When the data is complete and consistent in all respects, the next step is to filter the data according to our requirements. Since all the information is in one file and is jumbled, so separate the data of each semester with same attributes. The major attributes about the students (after data cleaning) are described in Table 5.1 and were used as basis for the research work.

Table 5.1. Attributes and their description

Attributes	Description
Exam code	The code of the exam, whether odd or even semester
Academic year	The year of studying the subject
Subject code	The unique code of the subject
Semester	The semester in which the student is studying
Subject	The subject of study
Enrollment no	Distinct roll number of the student
Student name	Name of the student
Grade	The grade obtained by the student, whether A,B,C,D or E

5.2 Description of workflow

The main steps of the workflow are as explained below:

1.Data fetching

Data fetching combines all the available data that can be used to resolve the data mining problem, into a set of instances. The data of our work is fetched from Thapar University students of Computer Science and Engineering department belonging to batch 2010-14. Each and every academic detail is taken for the efficient study of the approach.

2.Data Cleaning

The data cleaning process detects erroneous or irrelevant data and discards it. The data collected by us also had most common mistakes like inaccuracies, missing data and inconsistent data. Some of the attributes like tagging for, program code, section branch, semester type etc were irrelevant for our study. Therefore, they were cleaned from the data so that the main attribute could be used.

3.Data Filtering

Data filtering helps to reduce the large amount of information available to us. The most common types of filtering techniques are usually the selection of data subsets for educational data relevant to the intended reason. Our educational institute provided us a large amount of information about all the subjects and the grades attained accomplished by the students name and the enrolment number admitted in the course. However, we were only interested on a few subsets of courses or students depending on the proposed approach. For this reason, filtering is be used to select only a particular subset of desired data [46].

4.Data Transformation

Data transformation is the process of deriving new attributes from beforehand available attributes to assist a better interpretation of information. The data is separated according to the subjects and the grades obtained in those particular subjects. In addition, various formats were formed where the information is specific in accordance with the subjects, students, grade and7u semester.

5.3 Algorithms used

Now the data is ready for the data mining methods to be applied and generate the model so that it can be further used for analysis and prediction. The algorithms proposed are: SVM, Bayesian Network and Decision Tree (C5.0).

Support Vector Machines (SVM) is the newest technique for supervised machine learning. SVMs spin about the notion of the margin-any side of a hyper plane that splits two data classes [25].SVM algorithm aims to find the ideal hyper plane that provides largest minimum distance. Our study includes four kernels of SVM namely, RBF (Radial Basis Function), polynomial, sigmoid and linear. Bayesian Network is the directed acyclic graph (DAG) portraying dependence and independence between

variables. It consists of set of variables and the edges between those variables. Decision tree algorithm is a tree like framework that classifies the instances beginning from source node to the leaf node by electing the variables at each level so that the set of items are perfectly separated [33].In our study, the decision tree generated by C5.0 is used for the purpose of classification.

5.3.1 SVM

A support vector machine is a classification algorithm that supports text mining, nested data problems and pattern recognition. It is a supervised algorithm that can be used for classification and regression (binary and multi-class problem) anomaly detection (one class problem). It aims to find the optimal hyperplane which will maximize the margin between two classes. The word "support" is used because it is kind of supports the hyperplane on either side of the data points. Basically, it is used to generate the pattern from small datasets.

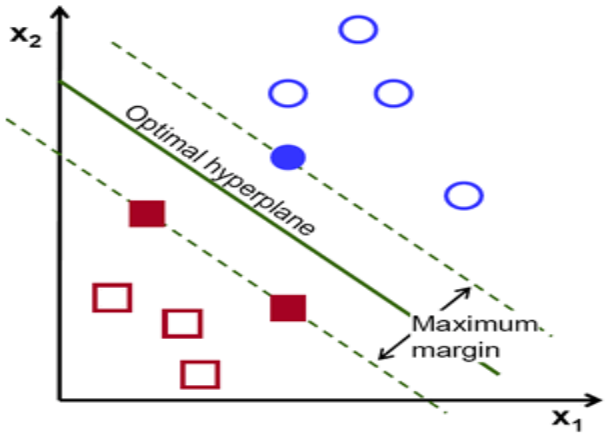


Fig. 5.2. Support vector machine[25]

Algorithm:

i. Designate an optimal hyperplane to maximize the margin
ii. Widen the above definition for non-linearly separable problems
iii. Map the data to high dimensional space where it is simple to classify with linear decision and reformulate problem so that data is mapped completely to this space.

5.3.2 C5.0

Decision tree is a classification technique of data mining that classify the instances having some featured values. In the decision tree, each node represents the feature in an instance and the branch represents the value.

Features of C5.0 [47]:

- Gives more accurate and efficient results
- Classify the result set with high accuracy and low memory usage
- Generates fewer rules and provides acknowledge about noisy and missing data
- Error rate is low and the prunes tree generates fast results
- Smaller decision tree is generated
- Performs feature selection and reduced error pruning techniques

5.3.3 Bayesian network

Bayesian network is a graphical model that depicts the probabilistic relation among the various attributes or instances [48]. The variables and their conditional dependencies are represented through directed acyclic graph (DAG). The nodes in the graph represent the random variable and the edges are the conditional dependencies.

5.4 SVM kernels:

1.Linear:

If the two classes are separated by a straight line, then those classes are known as linearly separable. The line can be separated by two critical members that define the channel, one for each class. These critical points are known as support vectors. Linear SVM scales linearly with the size of the training data set.

K (x, x') = $\langle x, x' \rangle$ ..... (1)

2.Sigmoid

Sigmoidal kernels are the hyperbolic functions. The origin of this kernel is from neural networks. The sigmoidal kernel is of the form:

K (x, x') = $\tanh(\gamma\langle x, x' \rangle + r)$. ..... (2)

3.Polynomial

Polynomial kernel in SVM represents the resemblance of vectors containing training samples in a feature space over polynomials of the original variables, therefore

allowing learning of non-linear SVM models. Polynomial kernel is quite useful in natural language processing but suffers from numerical instability. The polynomial kernel is of the form:

K (x, x') = $(\gamma\langle x, x'\rangle + r)^d$ ..... (3)

If d=1, it is a linear kernel where'd' is the degree and (x, x') are the input vectors.

If d=2, it is a quadratic kernel.

4.Radial Basis Function

The radial basis function is used to find set weights for a curve fitting problem. The learning helps to find out the surface in high dimensional space which provides best fit to the training data. The hidden layers supports a set of functions that comprises an arbitrary basis for input basis, such functions are known as radial basis functions. The RBF kernel is of the form:

K (x, x') = $\exp(-\gamma|x - x'|^2)$ ..... (4)

Where,$\gamma = 1/2\sigma^2$

σ is a free parameter

|x-x'| is the Euclidean distance

Chapter 6
Implementation and Results

Data mining is the extraction of information from large data sets and transformation of that information into some understandable structure for further use [49]. It is the process of selecting, exploring and modelling large amount of data by using different techniques to find useful patterns or models. Proper data mining technique should be available so that it can be used in many applications such as, social science, bank transactions, businesses, and psychology. In order to make this possible, proper data mining techniques should be available.

As the datasets grow in terms of complexity and size, some automated techniques like genetic algorithms, decision trees, SVM (Support Vector Machines), neural network comes into picture. Data mining exploits actual learning and discovers algorithms that are more efficient and allows such methods to be applied to larger data sets. Data mining process is applied with the intention of uncovering hidden patterns in large data warehouses [50]. The purpose of applying any data mining effort can be divided in two types: to generate descriptive models to solve problems and can be used to predict and solve problems.

6.1 Step-wise procedure

The data mining classification algorithms are different in many aspects such as: the learning rate, performance, speed, correctness, robustness, accuracy, etc [15].In this research, we examined thoroughly the impact of three algorithms for performance prediction: SVM, Bayesian network and C5.0 decision tree algorithm. The three classification techniques are employed to reveal the most appropriate way to measure student's performance.

Step 1:Semester-wise collect all the grades attained by each student in same sequence of subjects, as shown in Fig 6.1. The grades are collected as GRADES(i), where ‘i’ is semester and $1 \leq i \leq 4$.

$S_i = \{GS_1, GS_2, GS_3 ... GS_j\}$

Where,

GS=grade of the subject

S_i=semester, $1 \leq i \leq 4$

j=number of subjects corresponding to i^{th} semester. j=6 when *i* (i.e. semester) is 1 or 2 and j=7 when *i* is 3 or 4.

E.g.: S_2= {B, A, B, B, C, A} is the sequence of grades of the student in second semester having enrolment no. 101003010 and studying six subjects as highlighted in Fig. 6.1.

ENROLLMENTNO	GRADES(1)	GRADES(2)	GRADES(3)	GRADES(4)
101003001	A,D,B,D,D,C	D,B,B,D,C,D	C,C,C,C,E,E,D	E,C,B,B,D,C,D
101003003	C,B,B,B,C,B	C,C,C,C,C,C	B,B,B,C,C,D,B	B,C,A,B,C,B,A
101003004	D,C,C,B,B,C	D,D,E,D,D,C	B,B,C,C,C,C,D	C,C,C,C,B,D,B
101003005	C,D,B,C,C,B	D,C,D,B,C,B	C,D,D,B,B,C,C	A,B,C,B,C,B,C
101003006	B,B,A,A,B,A	A,A,B,B,A,A	A,B,A,A,A,A,B	B,A,B,A,B,A,A
101003007	C,C,E,C,D,D	C,D,D,D,F,D	D,C,E,D,D,E,E	D,D,D,D,C,C,C
101003008	B,B,B,B,B,B	B,B,B,C,B,B	A,B,C,B,B,B,B	A,C,B,B,C,B,B
101003009	C,D,D,C,D,B	D,D,F,D,D,D	C,D,D,D,C,E,D	C,D,C,E,E,C,E
101003010	A,A,B,B,A,A	B,A,B,B,C,A	A,A,B,A,A,B,B	A,B,B,B,B,A,C
101003012	B,C,D,D,C,C	D,D,D,F,D,C	C,D,D,D,C,C,C	D,C,E,C,C,D,D

Fig. 6.1. Collection of grades for four semesters

The grades and their corresponding performance criteria is shown in Table 6.1. This performance criteria is used for prediction in the final outcome i.e. OVERALLGRADE(F), where 'F' stands for final.

Step 2: Prepare the logic table (n^{th} level logic predicate) on the basis of GRADE in Table 6.1.

Level-wise logic order:

L_0: A – A → A

D – D → D

B – A → A

B – B → B

A – B → A

L_1: A – C → B

B – D → C

C -- E → D

L_2: A – D → B

B – E → C

D – A → C

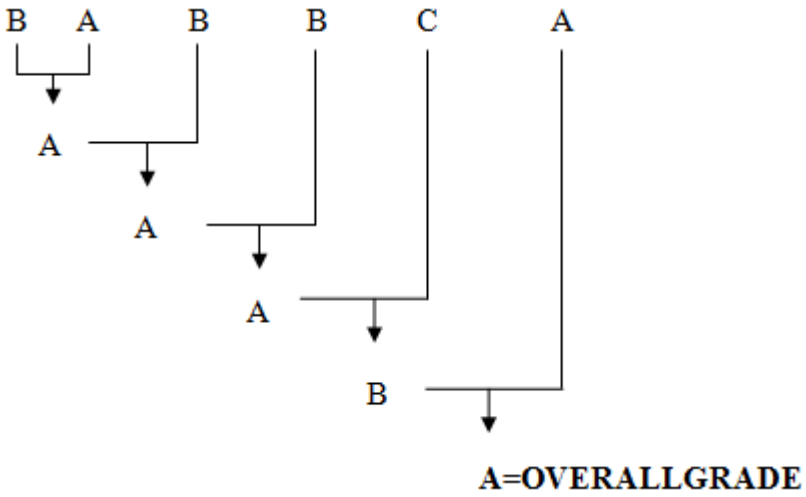
E – B → D

Table 6.1. Performance on the basis of grades

Grade	Performance	Marks
A	Excellent	9-10
B	Good	8-7
C	Average	6-5
D	Poor	4-3
E	Fail	2-1

L_3: A – E → C

E.g.: We have the sequence of grades {B, A, B, B, C, A} for the student having enrolment no. 101003010.While applying the step 2, take two consequent grades together and then compute the output in the way as shown below using Table 6.1. Like, B and A gives output A, then take this output as input for the next step. Now A and B gives output A. These steps are done using the above level-wise logic order until we reach to our final grade.



This shows that the performance of the student is **Excellent**, on the basis of Table 6.1. and is highlighted in Fig 6.2.

Step 3*:* Therefore, OVERALLGRADE(i), $1 \leq i \leq 4$ is computed for each semester(in the same manner as used above), where *i* is the semester. E.g. OVERALLGRADE(2) is the overall grade computed for second semester.

ENROLLMENTNO	GRADE(1)	OVERALL GRADE(1)	GRADE(2)	OVERALL GRADE(2)	GRADE(3)	OVERALL GRADE(3)	GRADE(4)	OVERALL GRADE(4)
101003001	A,D,B,D,D,C	C	D,B,B,D,C,D	C	C,C,C,C,E,E,D	D	E,C,B,B,D,C,D	C
101003003	C,B,B,B,C,B	B	C,C,C,C,C,C	C	B,B,B,C,C,D,B	B	B,C,A,B,C,B,A	A
101003004	D,C,C,B,B,C	B	D,D,E,D,D,C	C	B,B,C,C,C,C,D	C	C,C,C,C,B,D,B	B
101003005	C,D,B,C,C,B	B	D,C,D,B,C,B	B	C,D,D,B,B,C,C	B	A,B,C,B,C,B,C	B
101003006	B,B,A,A,B,A	A	A,A,B,B,A,A	A	A,B,A,A,A,A,B	A	B,A,B,A,B,A,A	A
101003007	C,C,E,C,D,D	C	C,D,D,D,F,D	D	D,C,E,D,D,E,E	D	D,D,D,D,C,C,C	C
101003008	B,B,B,B,B,B	B	B,B,B,C,B,B	B	A,B,C,B,B,B,B	B	A,C,B,B,C,B,B	B
101003009	C,D,D,C,D,B	B	D,D,F,D,D,D	D	C,D,D,D,C,E,D	D	C,D,C,E,E,C,E	C
101003010	A,A,B,B,A,A	A	B,A,B,B,C,A	A	A,A,B,A,A,B,B	A	A,B,B,B,B,A,C	B
101003012	B,C,D,D,C,C	C	D,D,D,F,D,C	C	C,D,D,D,C,C,C	C	D,C,E,C,C,D,D	C

Fig. 6.2. Overall grade computation for each semester

Step 4: Similarly, OVERALLGRADE(F) for each of the 126 student is computed as the final performance result. First ten results are shown in Fig. 6.3.

ENROLLMENTNO	OVERALL GRADE(1)	OVERALL GRADE(2)	OVERALL GRADE(3)	OVERALL GRADE(4)	OVERALL GRADE(F)
101003001	C	C	D	C	C
101003003	B	C	B	A	A
101003004	B	C	C	B	B
101003005	B	B	B	B	B
101003006	A	A	A	A	A
101003007	C	D	D	C	C
101003008	B	B	B	B	B
101003009	B	D	D	C	C
101003010	A	A	A	B	A
101003012	C	C	C	C	C

Fig. 6.3. Final performance computed using overall grades of four semesters

Step 5: Now various data mining methods are applied to OVERALLGRADE(F), which is the predicted performance of each student.

6.2 Experimental setup

The data for the model was collected for four semesters of Computer Science Engineering students, batch 2010-14 studying in Thapar University, Patiala. After eliminating the incomplete and unwanted data, the sample comprised 126 students having 'j' subjects. There are six subjects in first two semesters and seven subjects in semester 3 and 4.So,for semester1 and semester 2, j=6 and for semester 3 and semester 4, j=7.

Total number of records $=\sum_{j=1}^{i} \boldsymbol{S_j} * \boldsymbol{N}$

Total number of records for each student $=\sum_{j=1}^{i} \boldsymbol{S_j}$

Here, 'i' is the number of semesters. 'S_j' is the number of subjects corresponding to i^{th} semester and 'N' is the total number of students.

So, for 126 students there are 6*126 + 6*126 + 7*126 + 7*126 = 3276 data records. Each student is associated with 6+6+7+7=26 records. The outcome of each model is the student's predicted final result, which is then compared with our manual predicted performance result. The comparison of the performance results are analysed in terms of their accuracy and comprehensibility.

The three algorithms and their implementation using IBM SPSS Modeler are illustrated in the figures. There are four inputs: OVERALLGRADE(1) i.e. overall grade of first semester, OVERALLGRADE(2), OVERALLGRADE(3) and OVERALLGRADE(4) and the target being OVERALLGRADE(F).The type of all the input and output fields is nominal.

Fig. 6.4. represents the implementation of four kernels of SVM: RBF (Radial Basis Function), polynomial, sigmoid and linear. Each kernel elucidates the predictors in a different way so the output is taken in the form of tables. Fig. 6.5. shows how the decision tree algorithm, C5.0 is implemented and displays the rule set generated for overall performance using grades obtained in various semesters. The generated rules reveal the way in which the particular model computes the final performance when the overall grade of four semesters is provided as input. Fig. 6.6. is the decision tree obtained whose tree depth is 3 and root node is OVERALLGRADE(F). The other nodes are generated based on the various grades obtained in different semesters. Fig. 6.7. unveils the Bayesian Network algorithm implementation and the network model generated. This network briefly exhibits the conditional dependencies of the predictors with the target value, via directed acyclic graph.

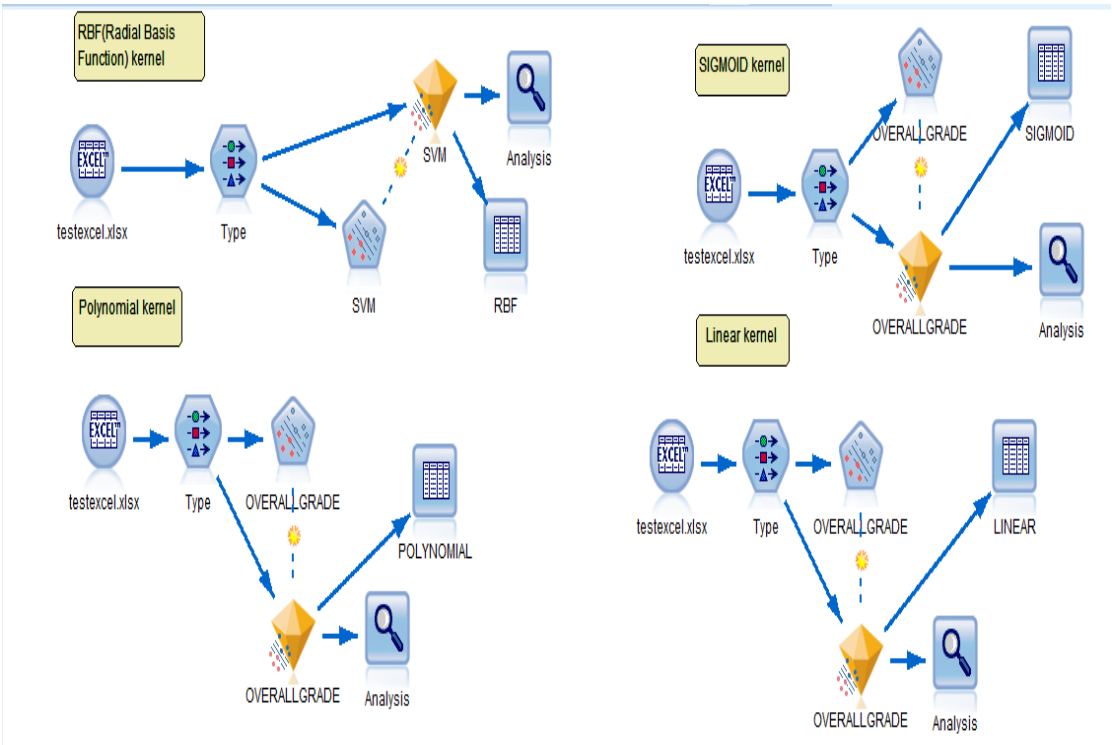


Fig. 6.4. Four kernels (RBF, polynomial, sigmoid and linear) implementation of SVM

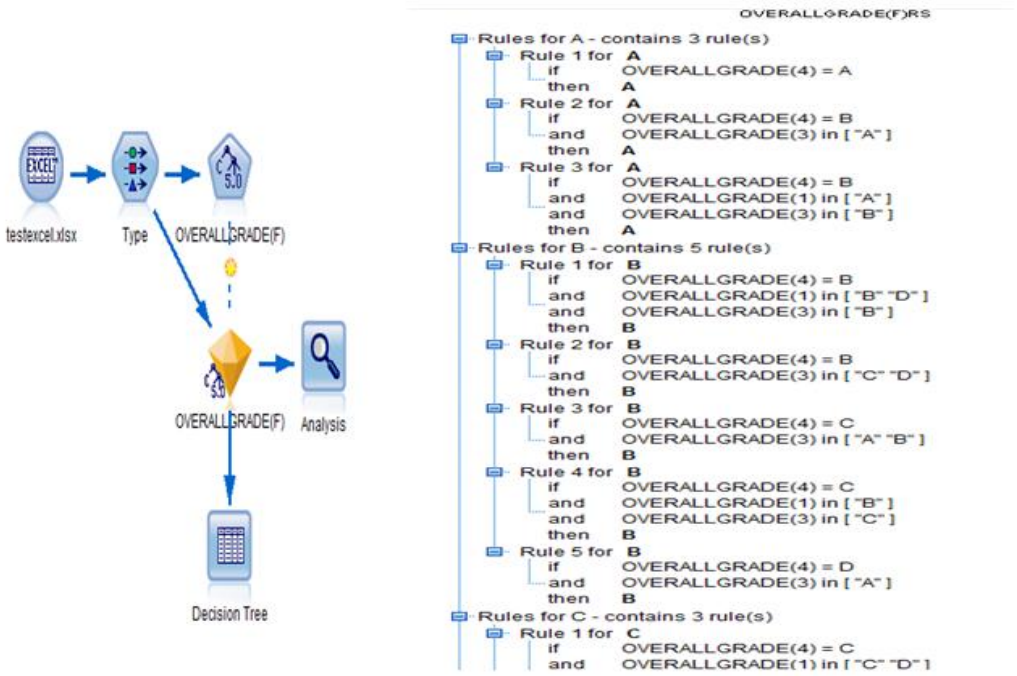


Fig. 6.5. Rule set generated by C5.0

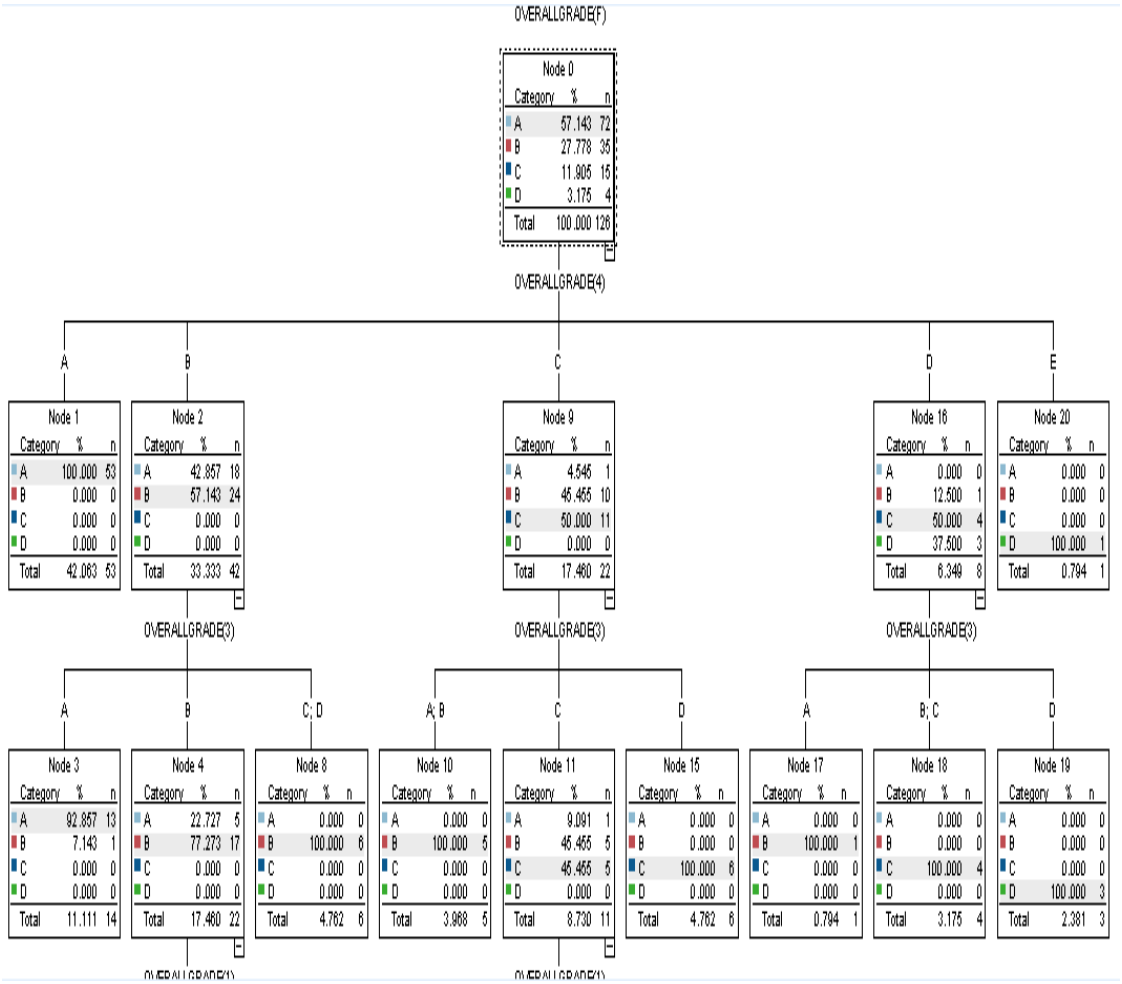


Fig. 6.6. Decision tree using C5.0

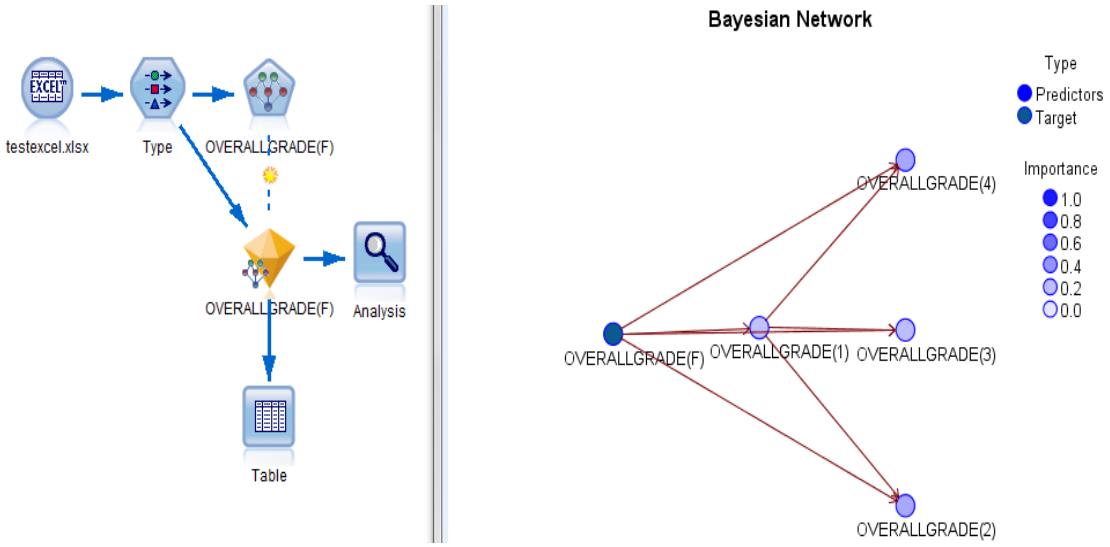


Fig. 6.7. Network model generation using Bayesian Network algorithm

6.3 Results and observations

The three algorithms provide different accuracy levels, i.e. each of them interprets the relevance of attributes in a different way. There are different evaluation criteria based on the classification algorithms used.

Table 6.2. compares the correct and incorrect instances obtained along with the prediction accuracy, when the three algorithms are applied.SVM has the highest accuracy as compared to other classifiers. The graph in Fig. 6.8. clearly shows the accuracy levels of the three algorithms, SVM being the most accurate. Similarly, the SVM kernels output is described in Table 6.3 in which two out of four kernels namely RBF and polynomial show same number of correct instances.

Table 6.2. Comparison of the three classifiers

EVALUATION CRITERIA	CLASSIFIERS		
	SVM	C5.0	Bayesian Network
Correctly classified instances	123	121	121
Wrongly classified instances	3	5	5
Prediction accuracy	97.62%	96.03%	96.03%

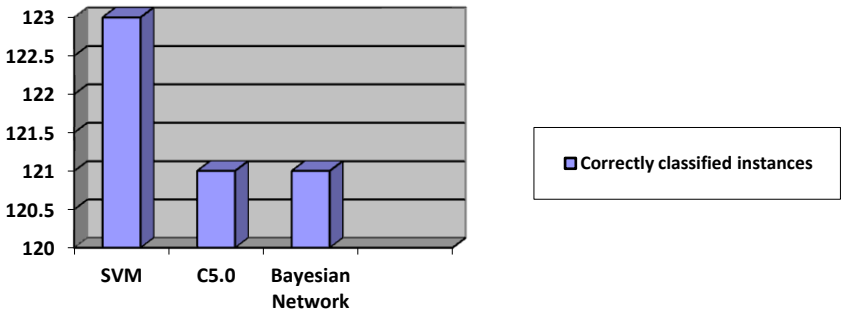


Fig. 6.8. Graphical representation of correctly classified instances

Table 6.3. Comparison of different kernels of SVM Model

EVALUATION CRITERIA	**SVM KERNELS**			
	RBF	Polynomial	Sigmoid	Linear
Correctly classified instances	123	123	80	120
Wrongly classified instances	3	3	46	6
Prediction accuracy	97.62%	97.62%	63.49%	95.24%

A report based on the confidence values is shown in Table 6.4. All the four kernels of SVM and C5.0 and Bayesian Network values are observed based on their predicted values, as generated by the respective models. Fig. 6.9. shows graphically the mean correct and mean incorrect values.

Table 6.4. Confidence values report

EVALUATION CRITERIA	**CLASSIFIERS**					
	SVM Kernels				C5.0	Bayesian Network
	RBF	Polynomial	Sigmoid	Linear		
Range	0.583 - 0.989	0.629-0.982	0.288-0.943	0.607-0.994	0.8-1.0	0.513-1.0

Mean correct	0.918	0.922	0.662	0.848	0.964	0.94
Mean incorrect	0.888	0.867	0.528	0.82	0.861	0.674
Always correct above	0.947 (38.89% of cases)	0.959(9.52% of cases)	0.646(28.57% of cases)	0.947(34.13% of cases)	0.941(62.7% of cases)	0.865(69.05% of cases)
Always incorrect below	0.583 (0% of cases)	0.629(0% of cases)	0.331(0.79% of cases)	0.607(0% of cases)	0.8(0% of cases)	0.598(1.59% of cases)

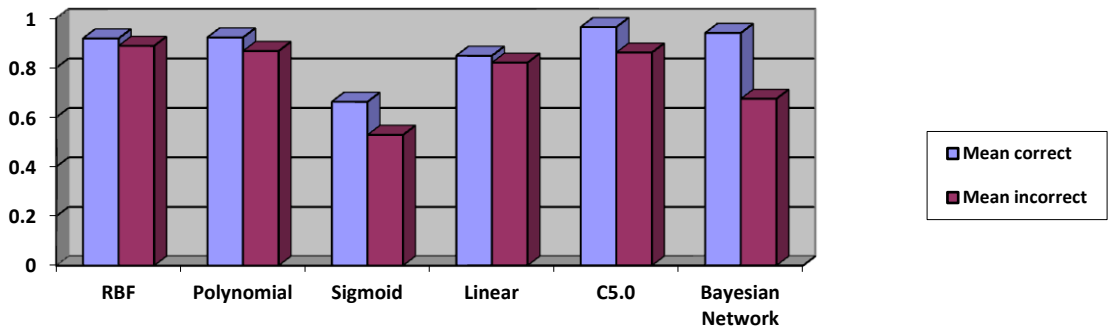


Fig. 6.9. Comparison of different algorithms based on mean correct and mean incorrect

Coincidence Matrix has proved to be good and easy enough to study the performance of the classifiers, statistically. Here, rows represent the actual value and columns represent the predicted value of the overall performance. From the Table 6.5. it is illustrated that SVM classifier produces better results. E.g. The first row of SVM represents that there are 70 students whose actual grade of final performance is A and predicted performance is also A. However, there are 2 students whose actual grade of performance is B but the tool predicts it to be A. And, there is no student whose actual performance is C or D but the prediction is A. The sum of each matrix is 126 i.e. the total number of students.

Table 6.5. Coincidence Matrix

CLASSIFIERS	A	B	C	D	AV / PV
C5.0	70	2	0	0	**A**
	2	32	1	0	**B**
	0	0	15	0	**C**
	0	0	0	4	**D**
SVM	70	2	0	0	**A**
	1	34	0	0	**B**
	0	0	15	0	**C**
	0	0	0	4	**D**
Bayesian Network	70	2	0	0	**A**
	1	34	0	0	**B**
	0	2	13	0	**C**
	0	0	0	4	**D**

(AV*:* Actual Value PV: Predicted Value A: Excellent B: Good C: Average D: Poor)

Chapter 7
Conclusion and Future Work

This work has been done on factual and real data. From the above analysis, we have concluded that SVM produces the best prediction results as compared to Bayesian network and C5.0. SVM exhibits higher accuracy i.e. 97.62%. The results indicate that RBF and polynomial SVM kernels perform better than sigmoid and linear. The proposed methodology can be adopted to help the teachers as well as the students to enhance the quality of learning and student's performance by taking significance decision at right time.

In future work, the study can be enhanced by including various demographic factors and more distinguishing attributes like SSC marks, HSC marks, projects undertaken etc. to obtain more accurate student performance and to determine student behaviour. Also, the work could be carried out with other modern techniques to acquire a wider approach and more reliable outputs.

References

[1] C. Romero, S. Ventura, “Educational data mining: a survey from 1995 to 2005”, *Expert Systems with Applications*, vol. 33, no.1, pp.135-146, 2007.

[2] D. P. Nithya, B. Umamaheswari, A. Umadevi, "A survey on educational data mining in field of education," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 1, pp. 69–78, Jan. 2016.

[3] E. Osmanbegović, M. Suljić, "Data mining approach for predicting student performance", *Economic Review – Journal of Economics and Business*, vol. 1, no. 1, pp. 3-12, 2012.

[4] R. Srivastava, M. Gendy, M. Narayana, Y. Arun, J. Singh, University of the future — “A thousand year old industry on the cusp of profound change”, Melbourne, Australia: Ernst & Young (Retrieved from http://www.ey.com/Publication/vwLUAssets/University of the future/$FILE/University of the future 2.12.pdf), 2012.

[5] Y. Ma, B. Liu, C. Wong, P. Yu, S. Lee, “Targeting the right students using data mining*”, In: Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 457–464, 2000.

[6] D. Suthers, K. Verbert, E. Duval, X. Ochoa, “Clow, MOOCs and the funnel of participation”, *In: International Conference on Learning Analytics and Knowledge*, pp. 185–189. ACM New York, 2013.

[7] Suthers, K. Verbert, E D. Silva, M. Vieira, “Using data warehouse and data mining resources for ongoing assessment in distance learning”, *In: IEEE International Conference on Advanced Learning Technologies*, pp. 40–45, 2002.

[8] J. Anderson, A. Corbett, K. Koedinger, “Cognitive tutors”, *J. Learn. Sci*., vol.4, no.2, pp.67–207, 1995.

[9] P. Brusilovsky, C. Peylo, “Adaptive and intelligent web-based educational systems”, *Int. J. Artif. Intell. Educ. 13*, vol.2, no.4, pp. 159–172, 2003.

[10] C. Romero, S. Ventura, E. Salcines, “Data mining in course management systems: moodle case study and tutorial”, *Comput. Educ.*, vol.51, no.1, pp. 368–384, 2008.

[11] V. Petrushin, "Multimedia Data Mining and Knowledge Discovery", J. Electron.

Imaging, vol. 17, no. 4**,** 2007.

[12] J. Srivastava, R. Cooley, M. Deshpande, P. Tan, "Web usage mining", *SIGKDD Explor. Newsl*., vol. 1, no. 2,pp 12,2000.

[13] I. Ognjanovic, D. Gasevic, S. Dawson, "Using institutional data to predict student course selections in higher education," *The Internet and Higher Education*, vol. 29, pp. 49–62, Apr. 2016.

[14] A. M. Shahiri, W. Husain, N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.

[15] E. Osmanbegović, M. Suljić, "Data mining approach for predicting student performance", *Economic Review – Journal of Economics and Business*, vol., no. 1, pp. 3-12, 2012.

[16] V. Ramesh, P. Parkav, K. Rama, "Predicting student performance: A statistical and data mining", *International Journal of Computer Applications*, vol. 63, no. 8, pp. 35-39, 2013.

[17] M. Goga, S. Kuyoro, N. Goga, "A Recommender for improving the student academic performance", *Procedia - Social and Behavioral Sciences*, vol. 180, pp. 1481–1488, May 2015.

[18] W. Xing, R. Guo, E. Petakovic, S. Goggins, "Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory", *Computers in Human Behaviour*, vol. 47, pp. 168–181, Jun. 2015.

[19] S. Natek, M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions", *Expert Systems with Applications*, vol. 41, no. 14, pp. 6400–6407, Oct. 2014.

[20] R. Campagni, D. Merlini, R. Sprugnoli, and M. C. Verri, "Data mining models for student careers," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5508–5521, Aug. 2015.

[21] C. Romero, S. Ventura, P.G. Espejo, C. Hervás, "Data Mining Algorithms to Classify Students", *In EDM*, pp. 8-17, June 2008.

[22] C. Antunes, "Acquiring Background Knowledge for Intelligent Tutoring Systems", *In EDM*, pp. 18-27, June 2008.

[23] M. Mavrikis, "Data-driven modelling of students' interactions in an ILE", *In EDM*, pp. 87-96, June 2008.

[24] H. Jeong, G. Biswas, “Mining student behavior models in learning-by-teaching environments”, *In EDM*, pp. 127-136, June 2008.

[25] K. Kaur and K. Kaur, "Analyzing the effect of difficulty level of a course on students performance prediction using data mining", *Next Generation Computing Technologies (NGCT), 2015 1st International Conference*, pp. 756-761, 2015.

[26] R. Asif, A. Merceron, and M. K. Pathan, "Predicting student academic performance at degree level: A case study", *International Journal of Intelligent Systems and Applications*, vol. 7, no. 1, pp. 49–61, Dec. 2014.

[27] D. Pelleg, A. Morre, “ X-means: extending K-means with efficient estimation of the number of clusters”, In *Proceedings of the Seventeenth International Conference on Machine Learning, San Francisco, CA, USA*, pp. 727-734, 2000.

[28] Harwati, A. P. Alfiani, and F. A. Wulandari, "Mapping student’s performance based on data mining approach (A case study)," *Agriculture and Agricultural Science Procedia*, vol. 3, pp. 173–177, 2015.

[29] J.M. Mativo and S. Huang, “Prediction of students' academic performance: Adapt a methodology of predictive modeling for a small sample size.” In *Frontiers in Education Conference (FIE), 2014 IEEE*, pp. 1-3, Oct. 2014.

[30] T. Mishra, D. Kumar & D.S.Gupta, “Mining Students’ Data for Performance Prediction.” In *Proceedings of International Conference on Advanced Computing & Communication Technologies*, pp. 255-263, 2014.

[31] E.P.I. García and P.M. Mora, “Model prediction of academic performance for first year students.” In *Artificial Intelligence (MICAI), 2011 10th Mexican International Conference on*, pp. 169-174, Nov. 2011.

[32] P.M. Arsad, N. Buniyamin & J.L.A Manan, “A neural network students' performance prediction model (NNSPPM).” In *Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on*, pp. 1-5, Nov. 2013.

[33] P. Guleria, N. Thakur, M. Sood, "Predicting student performance using decision tree classifiers and information gain," *Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference on*, Solan, pp. 126-129, 2014.

[34] V. P. Bresfelean, “Analysis and predictions on students' behavior using decision trees in weka environment,” *29th International Conference on information technology interfaces*, pp. 25-28, June 2007.

[35] K. Bunkar, U.K. Singh, B. Pandya & R. Bunkar, “Data mining: Prediction for performance improvement of graduate students using classification.” In *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*, pp. 1-5, Sept. 2012.

[36] K. Parmar, D. Vaghela & P. Sharma, “Performance prediction of students using distributed Data mining.” In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on*, pp. 1-5, March, 2015.

[37] K. Parmar, D. Vaghela & P. Sharma, “Prediction and Analysis of Student Performance using Distributed Data mining.” *International journal of emerging technologies and applications in engineering, technology and sciences (ij-eta-ets), Dec*. 2014.

[38] S. Huang & N. Fang, “Work in progress: Early prediction of students' academic performance in an introductory engineering course through different mathematical modeling techniques.” In *Frontiers in Education Conference (FIE), 2012* , pp. 1-2, Oct. 2012.

[39] P.A. Patil & R.V. Mane, “Prediction of Students Performance Using Frequent Pattern Tree.” In *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on* , pp. 1078-1082, Nov. 2014.

[40] K. Bunkar, U.K. Singh, B. Pandya & R. Bunkar, “Data mining: Prediction for performance improvement of graduate students using classification.” In *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*, pp. 1-5, Sept. 2012.

[41] "IBM SPSS Modeler," IBM, 2016. [Online]. Available: http://www-01.ibm.com/software/analytics/spss/products/modeler/. Accessed: Apr. 6, 2016.

[42] W. Klosgen, J. Zytkow, “Handbook of data mining and knowledge discovery”, *Oxford University Press, New York, 2002*.

[43] T. Mishra, D. Kumar & D.S.Gupta, “Mining Students’ Data for Performance Prediction.” In *Proceedings of International Conference on Advanced Computing & Communication Technologies,* pp. 255-263, 2014.

[44] K. J. Cios, W. Pedrycz, R. W. Swiniarski, L. A. Kurgan, “Data Mining: A Knowledge Discovery Approach”, *Springer, New York,* 2007.

[45] Association rules (in data mining) [Online] available:http://searchbusinessanalytics.techtar get.com/definition/association-

rules-in-datamining

[46] E. Mor and J. Minguillón, “E-learning personalization based on itineraries and long-term navigational behaviour”, *In: Thirteenth World Wide Web Conference, ACM, New York*, pp. 264–265, 2004.

[47] R. Pandya and J. Pandya, "C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning," *International Journal of Computer Applications*, vol. 117, no. 16, pp. 18–21, May 2015.

[48] D. Heckerman, “Bayesian networks for data mining.” *Data mining and knowledge discovery*, vol. 1, no.1, pp. 79-119, 1997.

[49] P. S. Pradnya, “Overview of predictive and descriptive data mining techniques”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5,no. 4, April-2015.

[50] C. Clifton, “Encyclopedia britannica: definition of data mining.” *Retrieved on*, vol. *9, no.*12, 2010.

List of Publications

1. Neha Choudhary and Ashutosh Mishra," Student Performance Measure By Using Different Classification Methods Of Data Mining ",*2016 International Conference on Advanced Communication, Control & Computing Technologies*, 2016.[Accepted]

YouTube Video link

The video link of the research work can be accessed from

https://youtu.be/Qm6pW432ZSA

Plagiarism Report