



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

A MACHINE LEARNING APPROACH FOR TRACKING AND PREDICTING STUDENT PERFORMANCE IN DEGREE PROGRAMS

**Dr.S.PRABHARAN¹, DAMERLA VEDAMUKHI², ABHESH KUMAR ROY³, D.ABHANUPRAKASH
REDDY⁴, BUDIGE RAVITEJA⁵**

ABSTRACT:

To choose applicants who will perform well academically at higher education institutions, an admissions system based on legitimate and reliable admissions criteria is crucial. This project focuses on strategies to help colleges make admissions decisions by predicting applicants' academic performance at universities using data mining tools. The suggested methodology was validated using data collected from 2,039 students enrolled in the Computer Science and Information College of a Saudi public institution between 2016 and 2019. The findings show that based on specific pre-admission factors, applicants' early university performance can be predicted before admission (high school grade average, Scholastic Achievement Admission Test score, and General Aptitude Test score). The outcomes also demonstrate that the score on the Scholastic Achievement Entrance Test is the pre-admission factor that most closely predicts future student success. Hence, admissions systems ought to give this score more weight. Also, we discovered that the Artificial Neural Network technique outperforms other classification techniques (Decision Trees, Support Vector Machines, and Naive Bayes) with an accuracy rate of over 79%.

INTRODUCTION

Universities should employ objective admissions criteria in order to find individuals who will succeed in their programmes. Also, prior to enrolling a student, each institution should estimate their potential using the most cutting-edge techniques. This would support university policymakers in creating efficient

admissions standards. To predict students' success, most universities struggle to analyse their enormous educational statistics [1]. This is due to the fact that they only employ conventional statistical methods, as opposed to more advanced and reliable prediction techniques like Educational Data Mining (EDM),

1 Assistant Professor, Dept of CSE, MALLA REDDY INSTITUTE OF ENGINEERING AND TECHNOLOGY(AUTONOMOUS), Dhulapally, Secundrabad, Hyderabad, Telangana, India.

2,3,4,5 UG Students, Dept of CSE, MALLA REDDY INSTITUTE OF ENGINEERING AND TECHNOLOGY(AUTONOMOUS), Dhulapally, Secundrabad, Hyderabad, Telangana, India.

which is the most popular technique for evaluating and forecasting student performance [2–6]. EDM includes initially obtaining relevant data and trends from a sizable educational database [2] in order to predict students' performance. Improved data enables more deliberate strategies to improve pupils' academic standing. By applying data mining techniques to provide more precise projections of applicants' future academic accomplishment, this research hopes to assist universities in their admissions procedures. This research contributes to the body of knowledge in many different domains. We first predict candidates' early academic success based on their profiles before they enrol using four data mining classification algorithms. The profile information that is most frequently used to forecast a student's success in higher education includes test and exam results, extracurricular activities, demographic information about the student body, cumulative grade point average, and social network relationships (e.g. [7]–[10]). Yet, the admissions process rarely takes into account elements that could be used to forecast student achievement, such as the results of pre-admission tests (e.g. [11]–[13]). These undervalued indicators are the main subject of this study. Also, we contrast four categorization techniques to see which one performs best in terms of accuracy, precision, recall, and F1-Measure while generating predictions about students' academic achievement. In order to

determine how the selection criteria for new students affect their GPA after the first semester, we next do a correlation coefficient analysis. To help decision-makers give that element more weight, we also identify the most accurate admissions criteria for student accomplishment. Lastly, the institution where this research was conducted changed its admissions criteria to provide varying weights to the characteristics that were determined to be the most crucial. This study proved the effectiveness of the new approach by comparing the cumulative grade point averages of freshmen accepted under the two different systems. The proportion of first-year students with exceptional or very good cumulative GPAs increased by 31%, while the proportion with passable or subpar GPAs decreased by 18%. This study stands out from others in the field of predicting student performance due to the large sample size of 2,039 students from the Faculty of Information and Computer Sciences (CCIS) at Prince Nourah bint Abdulrahman University (PNU) in Riyadh, Kingdom of Saudi Arabia (KSA). More female students attend this institution than any other in the entire world. The majority of earlier research in this field uses noticeably smaller samples to confirm the effectiveness of their models.

LITERATURE SURVEY

a novel system for analysing student performance that draws knowledge

from databases used in higher education, H. Guruler, A. Istanbulu, and M. Karahasan are the authors. Data mining is one method of knowledge discovery, which is used to identify significant patterns in vast amounts of data. Knowledge discovery software named MUSKUP has been created and tested on student data in order to investigate the variables influencing the success of university students. A decision tree classification is used as a data mining technique in this system. All of the tasks involved in the knowledge discovery process are kept together by this software system. The benefit of this strategy is that it gives users access to all of SQL Server's and Analysis Services' features through a single piece of software. The study was conducted using information from college students. The sorts of registration to the institution and the family's income levels were discovered to be related to student achievement, according the study's findings. (Includes 6 figures and 3 tables.)

Review of educational data mining S. K. Mohamad and Z. Tasir are the authors.

In the realm of education, data mining is tremendously helpful, particularly when analysing students' learning behaviours in an online learning environment. This is because data mining has the ability to analyse and reveal hidden information within the data itself, which would be challenging and time-consuming to perform manually. This review's objective is to examine the methods used by earlier

researchers to approach data mining and the most recent developments in this area of educational research. There are some research restrictions highlighted, and several potential study directions are offered.

The two most crucial factors in forecasting student accomplishment are attributes and prediction strategies. College students' cumulative grade point average (CGPA) is the single most crucial variable in predicting their academic performance. Its use has benefited numerous investigations (e.g. [7]–[10]). Academic performance studies also use grades from final exams, lab assignments, assessments, and quizzes (e.g. [8], [9]). Demographics, student activities, and social networks have been taken into consideration as independent factors in a very small number of studies. Pre-admission exams, for example, are rarely used in the admissions process to predict a student's success in college (e.g. [11]–[13]). We'll be looking at this in more detail. The success of students in their courses has been predicted using a variety of data mining classification techniques. For instance, in one study, ANN was used to forecast 505 students' performance in their eighth-semester coursework. Using limited information, a model was proposed to predict student achievement in particular courses using Decision Trees (32 and 42 students). a Naive Bayes analysis on the grades of 1,600 students in a single class. The study employs SVM on a dataset of

1,074 people to predict the academic success of freshmen who are at risk. Here, utilising a range of admissions characteristics as input quality, the prediction of first-year cumulative grade point averages in the computer engineering departments of Saudi public universities was explored. Few studies on the subject that were published in KSA were directed at institutions other than medical colleges (e.g. [11]–[13], [15]). EDM has not been used in these research, despite the fact that it might be able to reveal hidden patterns in large datasets. One of the rare studies on the subject that has been restricted to a computer science institution employed one of the EDM techniques (i.e., the J48 decision tree) to typically consider the ultimate GPA based on grades in all classes. The authors acquired data from the transcripts of 236 Saudi Arabian King Saud University (KSU) Computer Science College students. They identified the courses that had the greatest impact on the overall grade point average. To generate predictions regarding student performance, they only employed one EDM methodology on a small dataset, and they did not double-check their results using any other EDM methodologies.

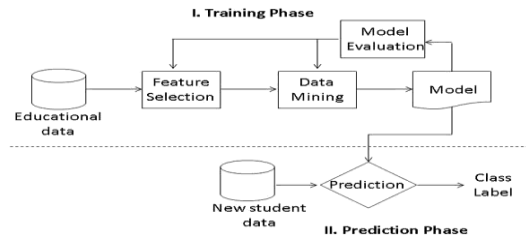
EXISTING SYSTEM:

To classify admitted applicants who will be accepted and not accepted, we are employing Decision Trees, Support Vector Machines, and Nave Bayes in existing

approaches. Analogrithms perform and perform more inaccurately. Finally, the university where this study was done opted to alter the weighting of its admissions criteria in light of the findings and suggestions made by this study. By contrasting the first-year CGPAs of new students admitted under the new method to those of students admitted under the old system, this study demonstrated the effectiveness of such a choice. The proportion of students with exceptional or very good first-year CGPAs rose by 31%, while the proportion of students with acceptable or subpar first-year CGPAs fell by 18%.

PROPOSED SYSTEM:

In this article, we suggest ANN methods that will improve model performance and reduce overfitting and underfitting. For the purpose of predicting student performance, a variety of data mining categorization algorithms have been used. For instance, ANN is employed in a study to forecast 505 students' academic achievement in their eighth semester. Using tiny student sample sizes, a study on Decision Trees suggested a model to forecast student performance in certain courses (32 and 42 students). A study on a group of 1,600 students uses naive Bayes to forecast achievement in a given subject. SVM is used in the study to forecast first-year performance for students at risk using a data set of 1,074 students.



Network of artificial neurons (ANN) Because it makes an effort to mimic how the human brain functions to tackle challenging problems, the artificial neural network (ANN) is a frequently used technique in evolutionary computational modelling (EDM). It is composed of a number of modules that each create a corresponding output after receiving a set of weighted inputs. Using ANN to predict student achievement has been the focus of various articles (e.g., [8]). As it can learn from a tiny sample and find all potential correlations between variables, we also used it. Moreover, ANN models outperformed conventional classification techniques in terms of how accurately they anticipated which candidates would be accepted and which would be rejected. The ANN model in this experiment uses a straightforward topology termed Multilayer Perception due to the small sample sizes (MLP).

Decision Tree, B. Nodes are rated in a decision tree from most important to least important. In the graph, each node stands for a property of an instance, and the edges between them reflect the possible ranges of values for that property. We use this technique since it is so widely used by scientists (e.g., [1], [6], [7], [9]). It

generates straightforward and understandable value projections. Moreover, it performs well with both categorical and numerical features, necessitates no laborious data preparation, and conveys rules that are user-friendly and clear [6].

VECTOR SUPPORT MACHINE, C (SVM) This strategy creates a hyperplane that divides things into their appropriate groups. A SVM method's generalisation error lowers as the hyper-distance plane's from the nearest object grows. SVM is employed in this study since it is effective for small datasets and has only been applied in a small number of prior research (e.g., [7]). It is also speedier than the alternative ways.

NAIVE BAYES, D. Naive Bayes is a simple probabilistic approach that applies Bayes' theorem by making independent assumptions about several variables. It computes the probabilities that each item falls into each group. This method was chosen for this inquiry due to its high computational efficiency, frequent use in pertinent literature, and overall ease of implementation. 3) To address the third research question in this work (Which datamining prediction approach works best in this study?), we employed accuracy, recall, precision, or F1-Measure metrics to analyse and compare the performances of the four models. (See Part A, Experimental Design below for a complete description of

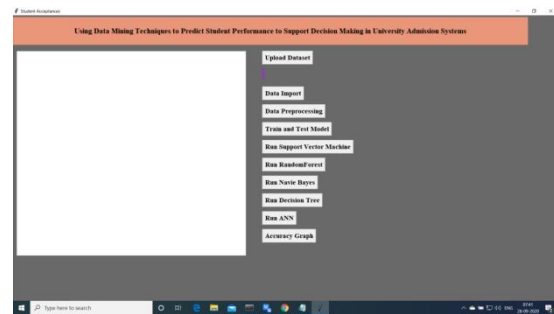
evaluation measures.) In order to address the fourth research question, we developed the same second stage of the study to compare the first-year CGPAs of incoming freshmen admitted in 2018-2019 using the new admission weights of criteria to the inaugural CGPAs of former students admitted in 2016–2 using the old admission weight of each criterion.

Working methodology:

For this investigation, information from the CCIS student database at PNU was used. Nonetheless, the strategies employed are universal and can be used in any university environment. These numbers are taken from the computerised academic database maintained by the Accepted & Registration Deanship. The required ethical clearance was provided by PNU's Institutional Review Panel (Number 19-0152). Data on 1,569 students from all three departments and two cohorts—902 students from the 2016–2017 school year and 667 students from the 2017–2018 school year—were collected in the first phase of our study. Second, we acquired 470 student records from the 2018–2019 academic year, when admissions were made using the updated weightings, from the three departments. Other student bodies experienced similar circumstances to those mentioned in the research's first section. We used this data to compare the first-year GPAs of newly accepted students to those of entering students under the prior weighting method. We constructed a category target variable

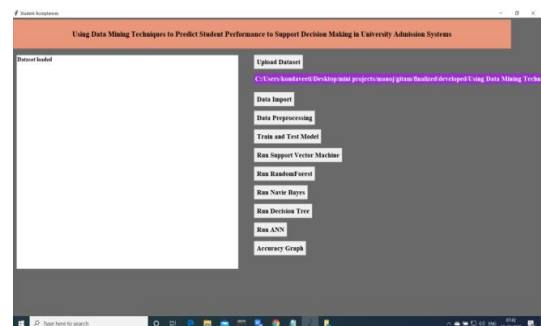
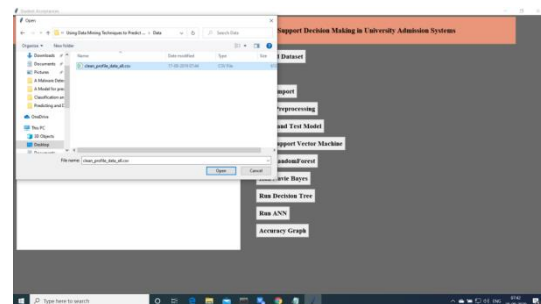
using the original numerical parameter, CGPA (class). The five-point PNU grading scale is split down into the following categories: outstanding (4.5), extremely good (3.75 to 4.5), great (2.75 to 3.75), ordinary (2.0 to 2.75), and terrible (2.0). An example of the grading system students used in the academic years 2016–17 and 2017–18.

Python final.py

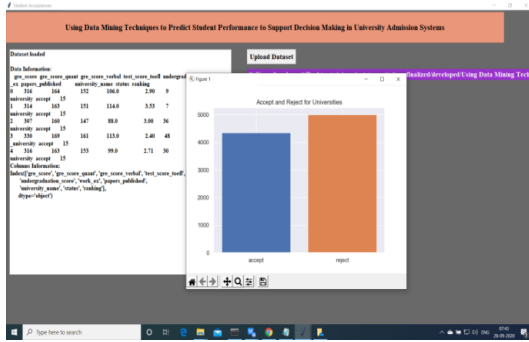


Above screen will be opened.

1. Now click on “Upload data ”

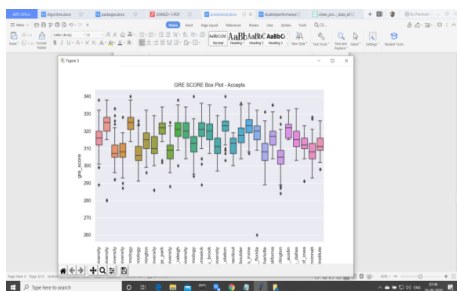
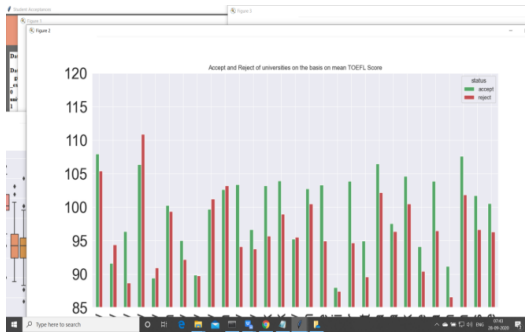


Import the data

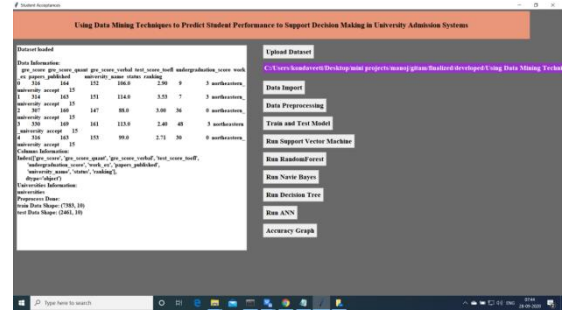


Upload the data and read the basic data information will be shown on the screen

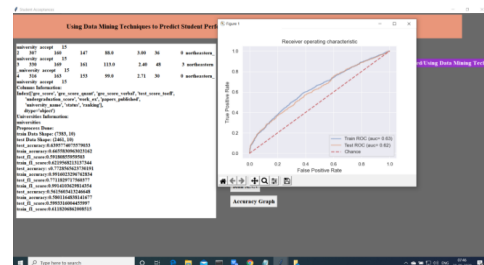
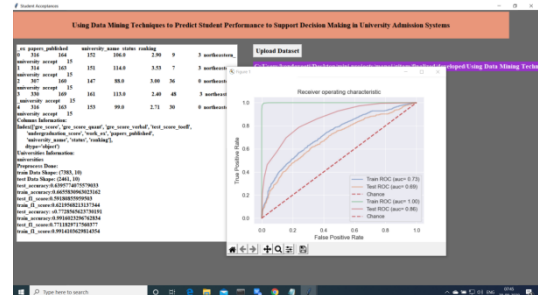
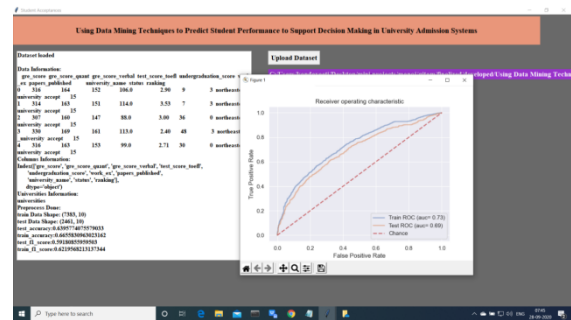
Data Preprocessing

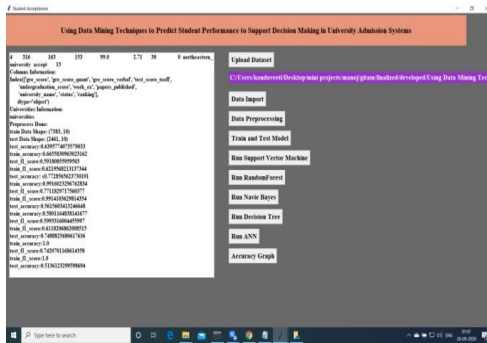
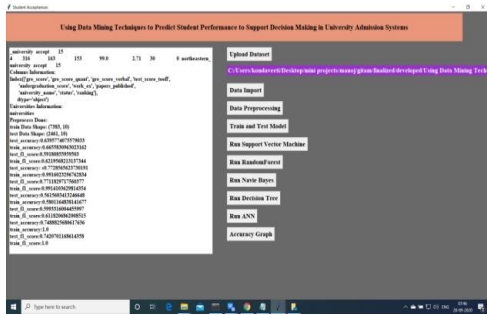


Now click on "Train and Test model". split the data into train and test and train will be used for training and to test the performance we are using test data

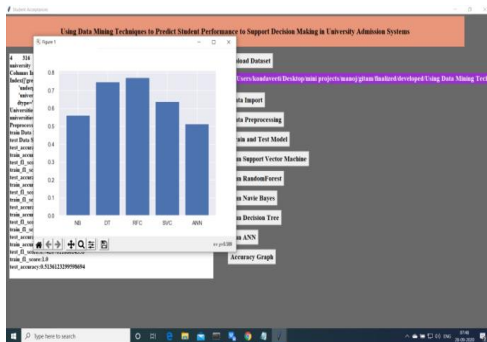


Now click on "Run Algorithms". Mentioned algorithms will be run on the data





Accuracy Comparison for all the models



Extension is random forest accuracy is increased.

CONCLUSION

This study's objective is to assist higher education institutions in making wise admissions decisions by assessing applicants' academic potential before granting admission. Four well-known data mining techniques—Artificial Neural Networks (ANN), Decision Trees, Support Vector Machines

(SVM), and Naive Bayes—were suggested and used to construct four prediction models. A dataset of 2,039 records of students enrolled in PNU, one of the biggest universities in KSA, served as the basis for the study. Nonetheless, the techniques employed are universal and can be applied in any higher education setting. According to the study, predictive modelling is beneficial at higher education institutions as decision-makers may use the models to plan and allocate the institutions' limited resources as efficiently as possible. The outcomes also demonstrate the possibility of creating a high performance model based on pre-admission data to forecast students' early success. For instance, the ANN model in this study was able to achieve performance accuracy levels of roughly 79.22%. The results of this study also demonstrate that the Decision Tree technique exceeds the others in recall and F1-Measure, while the ANN technique outperforms the others in both accuracy and precision metrics.

REFERANCES

[1] H. Guruler, A. Istanbulu, and M. Karahasan, "A new student performance analysing system using

knowledge discovery in higher educational databases,” *Comput. Edu.*, vol. 55, no. 1, pp. 247–254, Aug. 2010.

[2] S. K. Mohamad and Z. Tasir, “Educational data mining: A review,” *Procedia Social Behav. Sci.*, vol. 97, pp. 320–324, Nov. 2013.

[3] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1432–1462, Mar. 2014.

[4] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[5] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, “Educational data mining and learning analytics for 21st century higher education: A review and synthesis,” *Telematics Informat.*, vol. 37, pp. 13–49, Apr. 2019.

[6] C. Anuradha and T. Velmurugan, “A comparative analysis on the evaluation of classification algorithms in the prediction of students performance,” *Indian J. Sci. Technol.*, vol. 8, no. 15, pp. 974–6846, Jan. 2015.

[7] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, “Early segmentation of students according to their academic performance: A predictive modelling approach,” *Decis. Support Syst.*, vol. 115, pp. 36–51, Nov. 2018.

[8] M. Mayilvaganan and D. Kalpanadevi, “Comparison of classification techniques for predicting the performance of students academic environment,” in *Proc. Int. Conf. Commun. Netw. Technol.*, Sivakasi, India, Dec. 2014, pp. 113–118.

[9] S. Natek and M. Zwillling, “Student data mining solution–knowledge management system related to higher education institutions,” *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6400–6407, Oct. 2014.

[10] T. M. Christian and M. Ayub, “Exploration of classification using NB tree for predicting students’ performance,” in *Proc. Int. Conf. Data Softw. Eng. (ICODSE)*, Bandung, ID, USA, Nov. 2014, pp. 1–6.