Data-Driven Prediction of Student Performance to Support University Admissions

Mr. Potharaju Srinivas

Assistant Professor, Department of CSE,
Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India

ABSTRACT

To choose students who will do well academically in universities, it is crucial to have an admissions system based on legitimate and trustworthy admissions criteria. Methods for employing data mining to aid colleges in admissions decisions are the subject of this research estimate how well prospective students will do in college. For the purpose of testing the validity of the suggested technique, we utilized a dataset consisting of 2,039 student records from a Computer Science and Information College at a Saudi public institution between 2016 and 2019. The findings show that certain pre-admission parameters (high school grade average, Scholastic Achievement Admission Test score, and General Aptitude Test score) may predict how well a student would do in their first year of college. The findings also suggest that a student's score on the Scholastic Achievement Admission Test is the best predictive factor for admission. Thus, this score has to be given additional weight in selection procedures. We also discovered that the Artifical Neural Network method outperforms the other classication methods (Decision Trees, Support Vector Machines, and Naive Bayes) with an accuracy rate of over 79%.

INDEX TERMSKeywords: data mining methods, education data mining, predicting student success, admissions screening, and academic achievement.

INTRODUCTION

The admissions procedure at today's versities is difficult, but it is more difficult for computer and engineering schools. An effective admissions process is using tried-and-true methods of identifying and selecting students who are most likely to thrive academically within its curriculum. In addition, schools should use cutting-edge methods for gauging prospective students' potential academic success before accepting them. This would help policymakers at educational institutions establish effective admission standards.

However, most universities confront difficulties in analyzing their massive educational datasets to forecast students' performance [1]. This is because, instead of using effective prediction approaches like Educational Data Mining (EDM), the current trend in the field, they rely only on traditional statistical methods.

Fatih Emre Boran served as the associate editor responsible for organizing the evaluation of this article and giving final publication approval assess and foretell the academic progress of pupils [2] [6]. Extractive data mining (EDM) is the practice of mining relevant information and patterns from a large educational database [2] with the goal of predicting students' outcomes. Better data allows for more deliberate approaches to boosting students' academic outcomes.

This research aims to help colleges improve their admissions processes by using data mining tools to

assess candidates' potential contributions to the academic community before they are admitted.

The following are some of the specific issues probed by this research:

1) Can the admissions criteria that best predicts students' future success be determined? Is it feasible to use pre-admission exam results to foretell a student's early academic performance?

Which method of data mining prediction, from Artificial Neural Networks, Decision Trees, Support Vector Machines, and Naive Bayes, is the most accurate? results wise, that is.

How much of an impact does giving extra weight to the most significant entrance criteria have on first-year students' academic performance?

In various areas, this research adds to the existing body of knowledge.

To begin, we use data mining classication approaches to create four models that may predict applicants' early academic success based on their pre-admission proles. Quiz scores, nal exam scores, involvement in extracurricular activities, student demographics, total grades, and social network contacts (e.g. [7][10]) are some of the most common pieces of data used to predict student achievement in higher education that are included in these proles. Yet, factors that have been shown (e.g. [11][13]) to be reliable predictors of a student's academic success, such as test scores taken prior to enrollment, are seldom included in the admissions process. The current research is focused on these underutilized measures. Additionally, we determine the most effective of four classication strategies for predicting student performance with respect to accuracy, precision, recall, and F1-Measure measures.

Second, we conduct a correlation coefcient analysis to ascertain the link between admissions standards and freshmen CGPA. We also determine the best predictive admissions criteria in order to help decision-makers give greater weight to that factor.

Third, the institution where this research was done adopted a proposal to rebalance the criteria it uses to admit students. This research demonstrated the efficacy of such a choice by comparing the first-year CGPAs of new students accepted using the new method with those of new students admitted using the old approach. A 31% rise in the proportion of students with outstanding or very high first-year CGPAs and a 23% drop in those with acceptable or bad CGPAs may be attributed to these changes.

Finally, this research is distinct from others in the field of student performance prediction due to its huge sample size of 2,039 students from the College of Computer and Information Sciences (CCIS) at Princess Nourah bint Abdulrahman University (PNU), in Riyadh, Kingdom of Saudi Arabia (KSA). It has more female students than any other institution in the world. For model performance validation, most prior work in this area makes do with significantly lower sample numbers.

After scouring the academic archives of the Kingdom of Saudi Arabia (KSA), we were able to locate just one study [14] that analyzed the admissions criteria and their effect on student performance in the computer science and technology faculties of Saudi public institutions. In that research, the authors looked at how well students did on standardized tests leading up to college and how it predicted how well they would do in college. However, only one statistical model (regression) was used and no other major DM methods were compared to it. It's worth noting that

the research was conducted using data from only one division of the Computer College (IT). Furthermore, the institution where the research was performed did not implement the study's suggestions, therefore the study's validity was never put to the test.

To the best of our knowledge, the only previous research conducted on this issue in KSA (e.g. [11][13], [15], [17]) were limited to pre-medical students. In addition, rather of using EDM methods, statistical software was employed to analyze the data in these research.

Two research [18] in 2016 employed EDM to estimate students' GPAs based on their past performance in courses, but they did not investigate the correlation between students' academic records and admissions standards at universities.

In Section 2 we provide further information on the gaps in the linked literature and research.

The remaining sections of this work are laid out as follows.

In Section 2, we describe the gap between the existing body of work and the available research in this area.

Subject matter is defined under Section 3. The study's methodology is laid forth in Section 4. Results from the experiments are presented in Section 5. In Section 6, we provide our last thoughts, suggestions, and directions for further research.

II. RELATED WORK AND RESEARCH GAP

The quality of a university is most directly related to the success of its students. Presently, EDM is the method of choice for evaluators in the academic community and use it as a predictor of academic success because of the weight it carries in making choices [3, 7].

Attributes and prediction techniques are the two most important aspects in determining a student's final grade [19]. According to [19], the cumulative grade point average of college students is the single most utilized metric for forecasting their academic success. Numerous studies (e.g. [7][10], [20]) have made use of it. Assessments, quiz scores, labwork, and final exam grades are also often utilized by researchers as predictors of students' success in higher education (e.g. [8], [9], [21]). Extra curriculars, student demographics, and the web of interpersonal relationships have all been employed in a select few studies, although the majority of studies have focused on only one or two of (e.g. [19], [22]).

Pre-admission exams are one potential input variable that may be used to predict a student's success in higher education, although they are seldom employed (e.g. [11][13]) in the admissions process. This is what we'll be looking at in detail.

Predictions of academic achievement have been made using a variety of data mining classical methodologies. In research [21], for instance, ANN is used to foretell the eighth-semester academic achievement of 505 pupils. In regards to Decision Trees, one research [9] offered a methodology to predict student success in specific courses using small student sample sizes (32 and 42 students). Study [23] use Naive Bayes to forecast course grades for a sample of 1600 students. Using a dataset including information on 1,074 students, research [24] applies SVM to identify freshmen who are likely to underachieve.

NUMBER 8: 2020

The literature review shows that few research have investigated how entrance test results correlate with future academic success when applying to a college. This study fills a gap in the literature.

In this research, we utilized a variety of admissions factors as input variables to forecast first-year cumulative grade point averages (CGPAs) at a sample of Saudi Arabian public institutions offering degrees in computer science.

Most of the few published KSA-based research on the topic (e.g. [11][13], [15][17]) involve medical schools being duped. While EDM may help uncover hidden patterns in huge datasets, it has not been applied in these investigations. To forecast students' nal GPA based on grades in all courses, research [18] used one of the EDM methodologies (i.e., a J48 decision tree), making it one of the very few studies on this issue to be submitted to a computer science institution.

In order to compile their findings, the writers of [18] combed through the academic records of 236 Computer Science College students at King Saud University (KSU) in Saudi Arabia. They determined which classes had the most influence on the cumulative grade point average.

However, they only used a single EDM methodology on a tiny dataset to make predictions about student performance and did not double-check their work with any other EDM methods.

Related to this line of inquiry in KSA is research [25], which used three EDM methods (Naive Bayes, Artificial Neural Network, and Decision Tree) to make predictions about student performance based on past grades and forum activity. They concluded that Naive Bayes had the highest prediction accuracy (86%) out of the three methods by comparing their findings.

However, the emphasis of this research is on Saudi public university computer science students, and none of the previous studies [18], [25] have evaluated the connection between admission criteria and student performance.

As far as we are aware, study [14] is the only other published research on this issue. The authors of this research used regression analysis on data from 657 individuals to see whether taking a test before enrolling in school significantly affected their GPA. The final grade in high school was shown to have a greater impact on college GPA than did test results taken before to enrollment. The results of this investigation contradict these findings.

However, they only surveyed IT staff from KSU's multi-departmental Computer Science College.

In addition, they relied only on one statistical method (i.e., Regression Analysis) to make predictions about student performance, rather of using many EDM models.

III. UNDERSTANDING THE DOMAIN OF THE STUDY

The number of students enrolled in KSA's public universities' computer science and technology programs climbed from 2,959 in the 2004-2005 school year to 27,089 in the 2015-16 school year college/university term 2016-2017 [26]. The Saudi Ministry of Education [27] has standardized regulations and processes for student admissions to all Saudi public institutions since the early 2000s. Students applying to the computer science and technology undergraduate programs at Saudi public universities must first complete two standardized examinations offered by the National Assessment Center for Higher Education.

Both the Scholastic Achievement Admission Test (SAT) and the General Aptitude Test (GAT) fall within this category (SAAT). The GRE tests students' ability to understand and use quantitative and verbal information, whereas the SAAT evaluates their ability to do the same across five disciplines (biology, chemistry, physics, mathematics, and English) [28].

The High School Grade Point Average (HSGPA), the Scholastic Aptitude Test (SAT), and the Graduate Aptitude Test (GAT) are the primary factors used to determine admissions. Each public institution places a different value on each of these three factors, and each program sets its own cutoff for weighted averages according to its own annual quota of available slots. In the 2017–2018 school year, for instance, the three criteria (HSGA, SAAT, and GAT) each carried a different weight at the four most prestigious Saudi public universities: 30%, 40%, and 30% at King Saud University [29], 50%, 20%, and 30% at King Abdulaziz University [30], 20%, 50%, and 30% at King Fahd University of Petroleum & Minerals [31], and 60%, 20%, and 20% at King Abdulaziz University [32].

It is challenging to assign proper weights to these factors, however, since there are currently no studies in KSA that examine the connection between these characteristics and the performance of students in higher education. This research will aid in student selection by facilitating the use of a weighting method that gives due consideration to the most important factors.

IV. METHODOLOGY

The first research question we sought to address was, "Can the admission criterion that best forecasts an applicant's future academic success be determined?" To do so, we created a set of criteria that a model based on Linear Regression, a popular method for determining the connection between a set of predictor variables and a set of responder variables. We used the model to find out how the three entrance criteria (HSGA, SAAT, and GAT) relate to the dependent variable (CGPA during the first two semesters). Correlation coefficients are often employed in statistics to represent the strength of linear correlations between two variables, and we utilized them to characterize the link between each admission criterion and the CGPA. In addition, we employed the coefcient of determination to characterize the relative impact of each admission criterion on the students' rst-year CGPA.

2) Can we use pre-admission test results to foretell a candidate's early academic success and make admissions decisions based on these predictions?

We used four established data mining classication methods (Artificial Neural Network [ANN]), Decision Tree [DT], and Support Vector Machine [SVM]) to create four prediction models.

SVM) and Naive Bayes (NB).

Although there are many other methods for classifying DMs, the following four are the ones we settled on based on the salient traits we outline below.

ARTIFICIAL NEURAL NETWORK (ANN)

ANN is widely used in EDM because it is based on the idea that computers can learn to solve complicated problems in the same way that the human brain does. It is made up of a collection of parts that may take a weighted outputs based on a given set of inputs.

The use of ANN to forecast student performance has been the subject of a large number of articles (e.g., [8], [21], [23], [24], [33]). Since it can learn

from a small number of instances and can identify all conceivable relationships between variables, we utilized it as well. In addition, a recent research indicated that ANN models performed better than classication approaches at predicting which applicants would be approved and which would be rejected [33]. Due to the limited size of the datasets, we were able to employ a simpler ANN topology in this investigation: Multilayer Perception (MLP).

C. DECISION TREE

In a decision tree, the nodes are ranked from most important to least important. Nodes in the graph stand for characteristics of instances, while their connections show off the many ways in which those characteristics may be implemented.

This method was chosen since it values is simple and is thus popular amongst scientists (e.g., [1], [6], [7], [9], [34]).

It makes uncomplicated and easy-to-understand value predictions. Benefits include not requiring complicated data preparation, performing well with both category and numeric variables, and representing rules that may be simply understood and digested by users [6].

C. SUPPORT VECTOR MACHINE (SVM)

This method of classication creates a hyperplane that segments entities according to their classes. As the distance between the hyperplane and the closest object increases, the angle between the two decreases.the SVM method's generalization error worsens.

SVM is chosen in this work because it is well-suited for tiny datasets, and it has been utilized in a limited number of studies (e.g., [7], [23], [24]). The speed also makes it preferable to the alternatives [19].

D. NAIVE BAYES

Naive Bayes is a straightforward probabilistic method that uses independent assumptions across variables to apply Bayes' theorem. It calculates the odds of each possible outcome thing for every conceivable category. Because of its ease of use, high performance on practical problems, low computing cost, and widespread use in the relevant literature (e.g., [7], [23], [24], [35], [37]), we choose to implement this method in the present investigation.

3) We used accuracy, recall, precision, and F1-Measure metrics to analyze and compare the performance of the four models in order to answer the third question in this work (Which data mining prediction approach performs best in this study?). (For a detailed breakdown of evaluation metrics, please refer to Section A, Experimental Design below.)

4) To address the fourth research question (Do students admitted under the newweighting system achieve better academic performance during their first two semesters than past students admitted under the oldweighting system?), we designed the second stage of the study to compare the first-year cumulative grade point averages of students admitted for the 2018-2019 school year using the new admission criteria weights to those of students admitted in previous years using the old weights.

V. EXPERIMENTAL EVALUATION

A. EXPERIMENTAL SETUP

1) Dataset

CCIS at PNU student data was used for this analysis. However, the techniques used are open-ended in the sense that they may be implemented in any university setting.

The Admitted and Registration Deanship's office's computerized database served as the source for this information. Institutional Review Board at PNU provided the necessary ethical clearance (Number 19-0152). The Princess Noora Institution (PNU) is the biggest university in the world that is only for women, and it is the only public university in the Kingdom of Saudi Arabia that admits only female students. Nearly 50,000 students and over 5,000 faculty and staff members call this 32-million-square-foot campus home. The Computer and Information Science Center (CCIS) was founded in 2007 to meet the rising need for higher education in KSA for women in the fields of computer science and technology. Computer Science, Information Systems, and Information Technology are the three divisions that make up CCIS. Students spend four years completing the requirements of one of CCIS's three academic divisions. Two terms, or "semesters," make up a whole year.

The first phase of this research included collecting data from 902 students in the 2016–2017 academic year and 667 students in the 2017–2018 academic year across all three departments. The participants in this research were all women from Saudi Arabia (the central region of the Kingdom of Saudi Arabia).

Furthermore, all students do not work while college, but instead receive a government-provided monthly stipend. Three admission criteria—HSGA, SAAT, and GAT—were given 60%, 20%, and 20% respective weights in deciding which students from each cohort would be admitted.

PNU opted to alter the weighting of its admission criteria in light of the findings and suggestions from the rst stage of this research, which demonstrated that SAAT score was the entrance criterion that most correctly predicted student success. All of the weights have been adjusted to the 30%, 40%, and 30% breakdown for HSGA, SAAT, and GAT, respectively.

We gathered a total of 470 responses from participants in the study's second phase data from the 2018-2019 school year for all three departments, from which acceptances were calculated using the revised weights. Similar situations to those described in the rst stage of the research were encountered by other pupils.

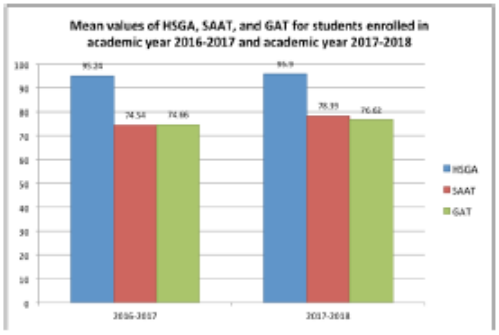

FIGURE 1. Mean values of HSGA, SAAT, and GAT for students enrolled in academic year 2016-2017 and academic year 2017-2018.

The first-year cumulative grade point averages of incoming students were compared to those of incoming students accepted under the previous weighting scheme using this data set.

Each student's record includes her CGPA after her first two semesters, as well as her scores on the three entry criteria (all numbers between 0 and 100). (all values between 0 and 5)

The Waikato Environment for Knowledge Analysis (WEKA) application [38] was utilized for the Data Mining approaches in this research, and hence the Excel document containing the data needed to be transformed into the standard format required by

WEKA. This is why we choose to adopt WEKA, a Java-based open-source program that has proven useful in the field of EDM research.

In the rst phase of the research, before using the EDM methods, we preprocessed the data.

We adopted measures including removing extraneous fields (such student ID, name, and address), spotting anomalies, getting rid of entries with missing information, and getting rid of duplicates. After the first data pre-processing, there were 1,430 student records left, including data for 828 students in the 2016-2017 school year and 602 students in the 2017-2018 school year.

Estimated mean values for HSGA, SAAT, and GAT are shown in Fig. 1.

Based on the initial numeric parameter CGPA, we built a category target variable (class). The ranges for each category are as follows: exceptional (4.5), very good (3.75 to 4.5), good (2.75 to 3.75), average (2.0 to 2.75), and bad (2.0) on the ve-level scale used in the PNU grading system. Figure 2 displays the cumulative grade point average distribution for both the 2016–17 and 2017–18 school years.

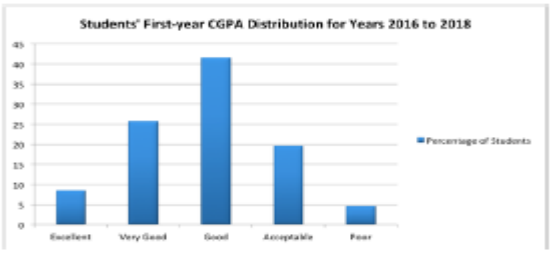

FIGURE 2. Students' first-year CGPA distribution for academic year 2016-2107 and academic year 2017-2018.

Similar to what we did with the data from 2016-2018, we performed some pre-processing on the data from 2018-2019. The Four hundred and thirty-seven student records were still available for the current school year (2018-2019).

Using four established data mining classication methods—Artificial Neural Network (ANN), Decision Tree, Support Vector Machine (SVM), and Naive Bayes—we constructed four prediction models. Ten-fold cross validation was used to construct each model; in this method, nine data sets were utilized for model training, and the tenth was used for model evaluation. This procedure was carried out ten times, once for each pair. The overall number of testable observations was increased because to this method. WEKA's default settings were used to run all models.

II. QUALIFICATIONS FOR REVIEW

Following these guidelines, we may evaluate how well the data mining models perform:

The percentage of cases for which a positive prediction was made (the True Positive Rate, or TP).

The percentage of times a result is wrongly labeled as positive (i.e., a false positive) is called the False Positive Rate (FP).

Number of occurrences when a negative prediction was made accurately; often called the True Negative Rate (TN).

Number of times a positive result was wrongly labeled as negative; also known as the False Negative Rate (FN).

Accuracy is the proportion of times an outcome is properly anticipated.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

Recall is the percentage of positives that are correctly predicted as positive and is measured with (2):

$$Recall = TP/(TP + FN) \quad (2)$$

Precision is the percentage of correct positive observations and is measured with (3):

$$Precision = TP/(TP + FP) \quad (3)$$

TABLE 1. The correlation coefficient between admission criteria and CGPA.

Independent Variables	CGPA
HSGA	0.20913
SAAT	0.44134
GAT	0.33783

TABLE 2. The coefficient of determination between admission criteria and CGPA.

Independent Variables	CGPA
HSGA	0.03523
SAAT	0.19257
GAT	0.13865

$F1$-Measure conveys the balance between the recall and the precision, and it emphasizes the performance of a classifier on common and rare categories, respectively [34]. It is measured with (4):

$$F1Measure = 2 \times Recall \times Precision/(Recall + Precision) \quad (4)$$

B. EXPERIMENTAL RESULTS AND DISCUSSION

1) MOST SIGNIFICANT CRITERION IDENTIFICATION

This study's findings that admission criteria may be used to predict students' first-year grade point averages (GPAs) highlight the significance of these factors as predictors of academic achievement at Saudi Arabia's universities.

Tables 1 and 2 show the results of the correlation and coefficient of determination between the factors, respectively, that we used to determine which admission criteria best predicts students' academic success after they arrive at university.

According to Table 1, SAAT scores are 44% more predictive of first-year CGPAs than any of the other independent factors.

Table 2 further demonstrates that the SAAT score has the biggest percentage influence on first-year CGPA of all of the admission factors, at 19%.

Therefore, the Scholastic Aptitude Test result is the most reliable indicator of future success in higher education.

For this reason, it merits more consideration. This finding is consistent with the findings of previous KSA-wide research (e.g., [11], [12], [15], [16]), which have shown that the SAAT score is the most significant positive predictor of student success in Saudi medical schools. One national research [14] suggests that the HSGA has a larger impact on students' GPAs than the other two factors combined, but our findings contradict this. There are a number of possible explanations for this disparity. One is that the HSGA score is meant to be a composite of the student's performance in all high school disciplines, not only those relevant to computer science. On top of that,

TABLE 3. Prediction techniques performance.

Classification Techniques	Accuracy	Recall	Precision	$F1$-Measure
ANN	79.22	78.03	81.44	79.70
Decision Tree	75.91	80.24	81.02	80.63
SVM	75.28	75.58	75.30	75.44
Naive Bayes	73.61	73.38	73.54	73.46

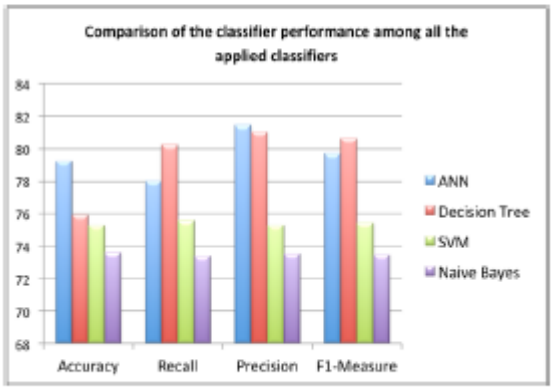

FIGURE 3. Comparison of the classifier performance among all the applied classifiers.

The curriculum at the secondary level is delivered in Arabic, whereas the college-level computer science classes are taught in English.

2) THE EFFICACY OF THE CLASSIFICATION METHOD

Table 3 displays the results obtained from using the aforementioned four classification methods.

In Fig. 3, we see a comparison of the classier performance of all the different classifiers that were used.

Applicants' early academic success at university may be predicted based on their entrance standards, as shown in Table 3. Each refined method yields excellent results in terms of accuracy, recall, precision, and F1-Measure. This suggests that these methods may be used by the relevant institution in order to assess prospective students' academic potential prior to admitting them.

Figure 3 demonstrates that the ANN classier approach excels in both the accuracy rate (79%) and the precision rate (81%), which are measures of classifier efficacy and predictive capability, respectively, compared to the other strategies studied. This confirms the findings of other research [8, 24]. Further, the DT classier approach exceeds the others in both recall (80%), which shows classier sensitivity, and F1-Measure rate (81%), which expresses the balance between recall and accuracy. This finding is consistent with those of previous investigations (e.g., [7], [35], [39]).

In addition, the SVM classier method provides reliable forecasts.

The NB classi er approach underperforms the others, consistent with previous research (see, for example, [36], [37]).

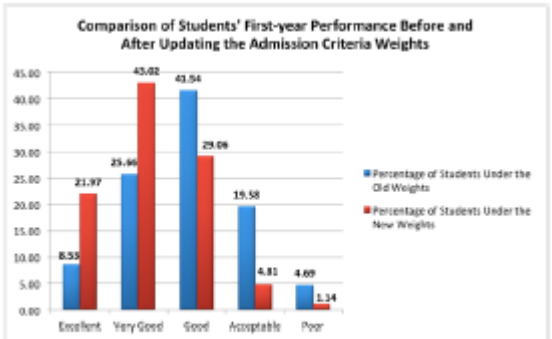

FIGURE 4. Comparison of students' first-year performance before and after updating admission criteria weights.

presenting the worst results across the board of the four key performance indicators.

These findings suggest that, in this case, When compared to other classifier methods, ANN classifiers tend to perform better across the board. There is substantial evidence in the literature that the ANN method is more accurate than other methods for predicting student achievement (e.g., [19], [24], [40]).

3) RAISE IN STUDENT PERFORMANCE

According to the study's findings and suggestions, the PNU Admitted and Registration Deanship has chosen to adjust the present admissions method by shifting the relative importance of various admissions criteria. This is done to improve the academic outcomes of first-year students. For the 2018-2019 school year, the SAAT criterion will be given 30% more weight than the HSGA criterion and 40% more weight than the GAT criterion. Prior to this change, the percentages were 60%, 20%, and 20%.

Graph 4 displays a comparison of first-year CGPAs from new students to those from prior years.

As can be seen in Fig. 4, the proportion of students with an outstanding or very high rstyear academic performance rose from 34% in 2016–2017 to 65% in 2018–2019 after extra weight was added to the most signicant entrance criteria (SAAT score). There was also a dramatic drop from the 24% who had an adequate or bad rst-year academic performance in 2016–2017 to the 6% who had such a showing in 2018–2019. This finding lends credence to the study's suggestion that the admission criteria best able to foretell students' future performance be given more weight.

A Final Thought

The purpose of this research was to aid admissions officers in making informed judgments about

which students to accept into higher education programs by assessing the likelihood that applicants would succeed once enrolled. Using four common data mining methods—Artificial Neural Network (ANN), Decision Tree, Support Vector Machine (SVM), and Naive Bayes—four different prediction models were suggested and built. The research used a dataset consisting of 2,039 student records from PNU, the biggest university in KSA. However, the approaches used are not exclusive to any one university but may be employed universally.

This research lends credence to the application of predictive modeling in higher education, where it may be put to good use by administrators in the distribution of scarce funds.

The findings also demonstrate the feasibility of developing a high-performance model to predict students' early success using data collected prior to enrollment. Example: the ANN model in this research achieved an accuracy rate of roughly 79.22% in terms of its performance.

This research also shows that the ANN method is superior to the others in terms of accuracy and precision, while the Decision Tree method is superior in terms of recall and F1-Measure. Results were poorest for Naive Bayes.

This research suggests that SAAT score should be given greater consideration as an admissions criteria since it is the most reliable predictor of future academic success. Decision-makers at PNU, where the research was performed, evaluated the admissions procedure after hearing about the findings and chose to give the Scholastic Aptitude Test score greater weight among the three entrance factors discussed.

This research compares the cumulative grade point averages of first-year PNU students who were accepted using the previous admissions weighted scheme with those who were admitted using the new approach. After implementing the new admission weighted method, 31% more students got outstanding or very high cumulative grade point averages (CGPAs) in their first year, whereas 18% fewer students achieved acceptable or bad CGPAs.

STATEMENTS OF SUGGESTIONS

Admissions officers at Saudi Arabian computer science universities should pay considerable thought to applicants' SAAT scores before admitting them.

Due to the fact that all engineering and computer science courses are taught in English, additional English-language proficiency tests should be a mandatory part of the application process for universities.

Factors like parental education, prior high school (national or international), social connection network, and programming expertise are all things that colleges and universities should think over before accepting a student.

Early identification of pupils in risky circumstances is made possible by the use of data mining tools in educational institutions will aid decision-makers in paying closer attention to them and developing a strategy to improve their academic standing.

FOR THE FUTURE

Data from a single Saudi Arabian institution was used across all classication methods. In order to verify

REFERENCES

[1] H. Guruler, A. Istanbullu, 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 classi_cation 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 classi_cation 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. Zwilling, ``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 classi_cation using NB tree for predicting students' performance," in *Proc. Int. Conf. Data Softw. Eng. (ICODSE)*, Bandung, ID, USA, Nov. 2014, pp. 1_6.

[11] J. Albishri, S. Aly, and Y. Alnemary, ``Admission criteria to Saudi medical schools. Which is the best predictor for successful achievement?" *Saudi Med. J.*, vol. 33, pp. 1222_1228, 2012.

[12] M. O. Al-Rukban, F. M. Munshi, H. Abdulghani, and I. Al-Hoqail, ``The ability of the pre-admission criteria to predict performance in a Saudi medical school," *Saudi Med. J.*, vol. 31, pp. 560_564, 2010.

[13] A. M. Alhadlaq, O. F. Alshammari, S. M. Alsager, K. A. F. Neel, and A. G. Mohamed, ``Ability of admissions criteria to predict early academic performance among students of health science colleges at King Saud University, Saudi Arabia," *J. Dental Educ.*, vol. 79, pp. 665_670, Jan. 2015.

[14] S. M. Hassan and M. S. Al-Razgan, ``Pre-university exams effect on students GPA: A case study in IT department," *Procedia Comput. Sci.*, vol. 82, pp. 127_131, 2016.

[15] M. F. Al-Qahtani and T. M. Alanzi, ``Comparisons of the predictive values of admission criteria for academic achievement among undergraduate students of health and non-health science professions: A longitudinal cohort

study," *Psychol. Res. Behav. Manage.*, vol. 12, pp. 1_6, Dec. 2018.

[16] A.-A. Dabaliz, S. Kaadan, M. M. Dabbagh, A. Barakat, M. A. Shareef, M. Al-Tannir, A. Obeidat, and A. Mohamed, ``Predictive validity of preadmission assessments on medical student performance," *Int. J. Med. Edu.*, vol. 8, pp. 408_413, Nov. 2017.

[17] A. A. Al-Ansari and M. M. A. El Tantawi, ``Predicting academic performance of dental students using perception of educational environment," *J. Dental Educ.*, vol. 79, pp. 337_344, Mar. 2015.

[18] M. A. Al-Barrak and M. Al-Razgan, ``Predicting students _nal GPA using decision trees: A case study," *Int. J. Inf. Edu. Technol.*, vol. 6, no. 7, pp. 528_533, 2016.

[19] A. M. Shahiri, W. Husain, and N. A. Rashid, ``A review on predicting Student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414_422, Jan. 2015.

[20] N. Thi Ngoc Hien and P. Haddawy, ``A decision support system for evaluating international student applications," in *Proc. 37th Annu. Frontiers Edu. Conf. Global Eng., Knowl. Borders, Opportunities Passports*, Milwaukee, WI, USA, Oct. 2007, pp. F2A-1_F2A-6.

[21] P. M. Arsad, N. Buniyamin, and J.-L.-A. Manan, ``A neural network students' performance prediction model (NNSPPM)," in *Proc. IEEE Int. Conf. Smart Instrum., Meas. Appl. (ICSIMA)*, Kuala Lumpur, Malaysia, Nov. 2013, pp. 1_5.

[22] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, ``Predicting students' _nal performance from participation in on-line discussion forums," *Com- put. Edu.*, vol. 68, pp. 458_472, Oct. 2013.

[23] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, ``Models for early prediction of at-risk students in a course using standards-based grading," *Comput. Edu.*, vol. 103, pp. 1_15, Dec. 2016.

[24] G. Gray, C. McGuinness, and P. Owende, ``An application of classi_cation models to predict learner progression in tertiary education," in *Proc. IEEE Int. Advance Comput. Conf. (IACC)*, Gurgaon, Indian, Feb. 2014, pp. 549_554.

[25] A. Mueen, B. Zafar, and U. Manzoor, ``Modeling and predicting students' academic performance using data mining techniques," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, pp. 36_42, Nov. 2016.

[26] Ministry of Education in Kingdom of Saudi Arabia. (Mar. 18, 2019). *Ministry Deputyship for Planning and Development*. [Online]. Available: https://departments.moe.gov.sa/PlanningDevelopment/RelatedDepartments/Educationstatisticscenter/EducationDetailedReports/ Pages/default.aspx

[27] National Center for Assessment (Qiyas). (Mar. 18, 2019). *Establishment of National Center for Assessment in Higher Education*. [Online]. Available:http://www.qiyas.sa/en/About/Pages/establishment.aspx

[28] National Center for Assessment in Higher Education. (Mar. 18, 2019). *Education Exams*. [Online]. Available: http://www.qiyas.sa/en/pages/default.aspx

[29] King Saud University. (Mar. 19, 2019). *Deanship of Admission and Regis- tration*. [Online]. Available: https://dar.ksu.edu.sa/en/e-admission

[30] King Abdulaziz University. (Mar. 19, 2019). *Admission*. [Online]. Available: http://www.kau.edu.sa/admission.aspx