Machine Learning Techniques for Predicting Student Performance

Rushali Deshmukh [#1], Atharva Kulkarni [*2], Aditya Kumthekar [#3], Prasanna Kottur [#4], Soham Raut[#5].

JSPM's Rajarshi Shahu College of Engineering[12345]

radesh19@gmail.com[#1], atharvaakulkarnii@gmail.com[*2], adkumthekar@gmail.com[#3], prasannakottur2@gmail.com[#4], rraut0700@gmail.com[#5]

Article Info

Page Number: 168 - 189

Publication Issue:

Vol 71 No. 4 (2022)

Article History

***Article Received**: 25 March 2022*

***Revised**: 30 April 2022*

***Accepted**: 15 June 2022*

***Publication**: 19 August 2022*

Abstract

Predicting pupils' performance has become increasingly tough because of enormous data. Recently, there has been a lack of a framework to examine and monitor student growth and performance in India. The following are the key causes for this. To begin with, the previous existing prediction methods are not sufficient to determine the appropriate approaches for student performance prediction in Indian institutions. Another reason is a dearth of research on the elements that influence students' progress in specific courses within the course setting. As a result, there is a comprehensive literature review on applying data mining approaches to predict student performance. It is advocated that pupils' academic achievements be improved. The purpose is to give a recap of data mining approaches that were used to predict student performance. This algorithm also helps to find the most important characteristics in a student's data. Utilizing educational data mining approaches could boost students' success and achievement cost-effectively. Students, educators, and academic institutions would benefit and be impacted due to the features of a student's data.

Keywords: - One Hot Encoding, Data Extraction, Neural network, SVM.

---

1. Introduction:

Student performance is critical in advanced education organizations. This is because one of the requirements for an excellent university is a strong academic record [1]. There are numerous definitions available. Based on the prior literature, students' performance was evaluated. According to Usamah et al. (2013), student performance can be improved. The learning valuation and co-curriculum were measured [2]. The majority of the research, on the other hand, mentioned that approximately Graduation is used to determine a student's success. In general, most Indian higher education institutions employed the last results to assess pupils' performance. Concluding grades are determined by the course arrangement, evaluation mark, last test result, and additional activity [2]. It is critical to evaluate pupils' progress and the efficiency of the learning procedure. A strategic program can be well prepared during a student's time in an institution by analyzing their performance [3]. Many strategies for evaluating student performance are now being proposed. One of the most often used strategies for analyzing student performance is data mining. Data mining has recently become popular in the educational field [4]. It's referred to as instructive data mining. Edu educational data mining is a method for obtaining beneficial info and designs from a vast database of instructive information. [5]. Students' performance can be predicted. As a result, instructors will be better able to provide an effective instructional strategy. Educators could also keep track of their pupils' progress. Students' learning activities could be improved, allowing the management to increase the system's performance. As a result, data mining techniques can be used for unique purposes with various entities. Machine learning systems offer much potential for assisting instructors in detecting poor student performance by providing an early warning system. As a result, instructors can devote more time to such struggling pupils to prepare them for summative tests. We used different machine learning algorithms to determine the prediction accuracy on the historical results of a course given in a bachelor's in computer information systems program. These models will be utilized to utilize students' formative exams, and if the model forecasts that a pupil is more likely to fail a course, alternative educational tactics will be applied to improve his or her learning experience. The following part discusses relevant work, followed by a problem statement, an explanation of the experiment specifics and findings, and finally, a conclusion.

The proposed systematic review's goal is to support the study's goals, which are as follows:

1. To look at the faults in present prediction methods and pinpoint them.
2. To investigate and determine the variables used to assess student achievement.
3. This study aimed to look into factors that predict student success.

The following part will present the survey approach for determining student achievement. Then, in Section 3, there will be a debate on research questions. Next, the outcomes of the existing prediction algorithms are presented in-depth in Section-depth ally, Section 5 discusses the conclusion and future work.

2. Methodology

A systematic relational review aims to find relevant approaches for current parameters, address research gaps, and lay down new research efforts in the proper context [6]. An organized review of the contemporary literature is intended to hold up the research topics that have been proposed. In addition, the research questions that will guide the outcomes will be identified. This is also helpful in deciding the scope of the investigation.

2.1. Research Issues

Analyzing the current study of student performance prediction necessitates the use of appropriate research questions. Table 1 lists the requirements for research survey questions on Kitchenham's approach to formulating research questions, including PIOC i.e. Population, Intervention, Outcome, and Context [6][7].

Table 1: Research Measure

MEASURES	**DETAILS**
POPULACE	School/University (Student performance)
INVOLVEMENT	Methods/Way’s for a forecast
RESULT	Successful forecast technique
CIRCUMSTANCE	All institutions

As a result, the following research questions are offered in this learning:

- Q1: What are the most significant elements for forecasting student performance?
- Q2: What approaches are used to predict student performance?

Nevertheless, before going deeper into this research, it is best to start with pilot learning. The goal of the trial learning is to see if the study questions are appropriate for the goals of the study. The research will next go into how to conduct a pilot study using a search approach.

2.2. Exploration Methodology

In an orderly analysis, a well-organized search plan is essential for locating all applicable pieces of work in the exploration results. As a result, a thorough exploration of study articles was led to address the study's specific research questions. Steps were taken by Kitchenham et al. (2010) [7]. To build the search words utilized in this systematic review, the following steps were taken. The following are the search terms that came up as a result: (student performance) AND (educational data mining) AND (systems OR application OR method OR process OR system OR technique OR methodology OR procedure) AND (systems OR application OR method OR process OR system OR technique OR methodology OR procedure) AND (prediction, estimation, or evaluation) The following choices are made as part of the search strategy:

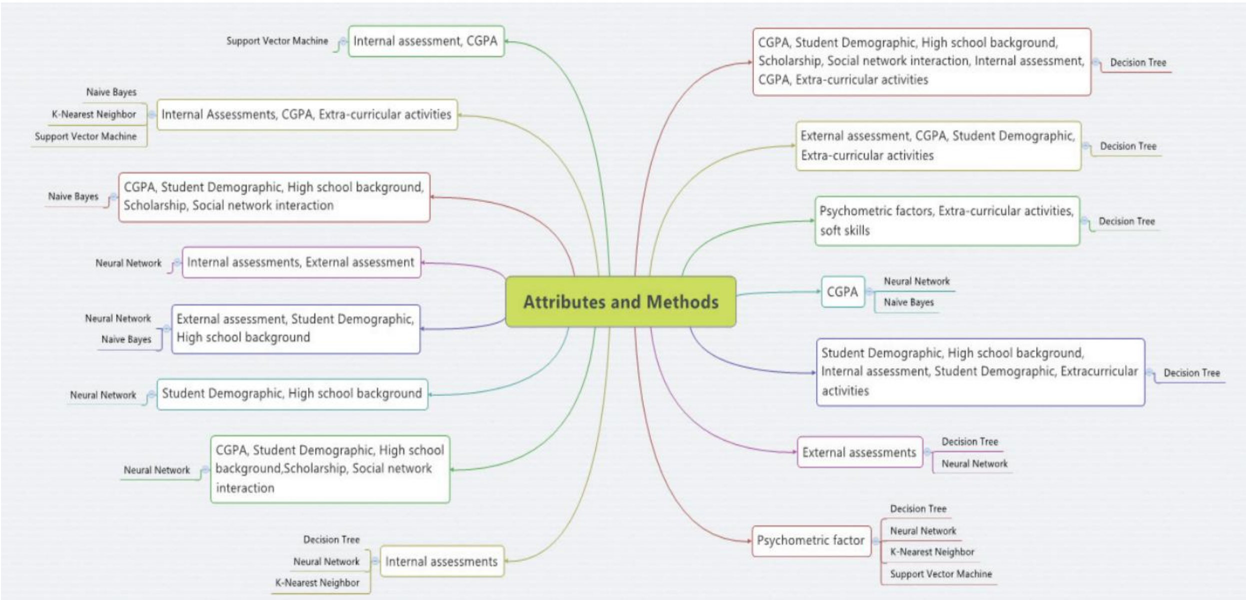

Fig. 1 There are a lot of similar features and approaches for predicting a student's performance.

3. Related Work

This section delves into the important factors predicting a student's success. The two main components of forecasting a student's success are attributes and prediction algorithms. A graphical illustration of a collection of popular criteria and approaches for anticipating student achievement can be seen in Figure 1. The first section will focus on crucial features that can predict student success, while the second section will focus on prediction approaches.

3.1. Key characteristics that are utilized to predict student achievement

The systematic literature review aims to identify critical factors that can be utilized to predict student achievement.

The cumulative grade point average (CGPA) and inner valuation are two frequently used metrics. Ten out of thirty studies were chosen. [5–8, 9–10, 3–11, 12–13, 14–15, 16] CGPA was one of the most important factors in predicting student achievement. Most researchers use CGPA since it has a measurable price in upcoming learning and occupational flexibility. It can be interpreted as evidence of an abstract promise that has been realized. [2].

According to the coefficient correlation analysis [3], CGPA is an extremely substantial input variable by 0.87 when likened to further components. In addition, according to Christian and Ayub's research [14], One of the most crucial elements in defining whether or not pupils will be able to finish their education is their CGPA. In this study, interior valuation was labeled such as. Task grades, tests, workshops, session assessments, and occurrences were used to define internal evaluation in this study. All traits will be referred to as interior valuation. Examiners frequently use the traits to forecast student achievement. [5, 17, 18, 19, 20, 21, 10, 22, 23, 12].

The demographics of the students and external assessments are the next most regularly used attributes. Students differ in sex, stage, family history, and incapacity. [2, 18, 9, 3, 24, 11, 25, 13, 14].

While external evaluations are important, Amirah Mohamed Shahiri et al. as a final test grade for a specific subject. [5, 17, 19, 26, 27, 24, 28, 13, 29]. Most researchers have employed student demographics such as gender since they had different approaches to women and man pupils in their learning procedure. [2]. Meit et al. (2007) discovered that, in comparison to male students, most female students have a variety of favourable learning styles and habits. [30]. Female students are more self-directed, well-organized, and concentrated in their academics than male students. In their studies, female students, on the other hand, employ effective learning strategies [31].

They were able to take advantage of self-motivation, organization, and rehearsing. As a result, one of the most critical elements impacting a student's performance has been identified as gender.

Extracurricular doings [5, 18, 12, 13, 32], college history [9, 24, 11, 25], and social interaction network [9, 33, 26, 21, 34] are the three most commonly used factors in predicting students' success. Each of these characteristics was employed in five out of thirty investigations.

Another study found that various researchers employed psychometric factors to predict student success [35, 36, 32, 37]. Psychometric characteristics such as pupil curiosity, educating behavior, engagement period, and family funding have been identified. They have taken advantage of these traits to design a system that appears to be very strong, humble, and comprehensible. In addition, it allows the speaker to calculate pupils' progress based on their interests and habits [12]. However, according to many studies, these qualities are rarely employed in predicting student performance since they focus on qualitative data, which is difficult to gather from defendants.

3.2. The strategies for predicting student performance that has been utilized

Predictive modeling is frequently used in educational data mining methodologies to forecast student achievement. Classification, regression, and categorization are all part of predictive modelling. Classification is the most common task used to predict student progress. Several strategies for predicting student performance on the classification task have been tried. Decision trees, Artificial Neural Networks, Naive Bayes, K-Nearest Neighbour, and Support Vector Machines are among the methods employed. The following section will investigate how data mining tactics organized by algorithms can be used to predict student achievement.

3.2.1. Tree of Choice

One of the most commonly used prediction tools is the Decision Tree. Because of its simplicity and comprehensibility, the majority of academics have used this technique to disclose minor or huge information structures and forecast rates. [8, 9, 13].

Decision tree models, according to et al. (2008), are easy to understand and can be used to solve problems because of their reasoning process. [22] Converted directly into a set of IF-THEN rules. There are about ten (10) papers, as stated in Table 2,

that have used the Decision Tree as a tool to assess student achievement. Previous studies that have used The Decision Tree approach is used to predict dropout factors in student data for academic performance [8], as well as predict academic performance.

MCA students' third-semester performance [32], as well as estimating a student's appropriate career path based on their grades.

[18] Behavioural patterns of the pupils' performance are evaluated using topographical extraction from recorded data in an Excel spreadsheet.

A web-based method for schooling. Pupils' last result [23], Last increasing grade point are instances of datasets.
Average (CGPA) [3] and grades earned in specific classes [22]. All of these datasets were examined and evaluated to discover
the key characteristics or factors that may influence a student's performance [28, 13]. Then there's the appropriate data mining.

The use of an algorithm to predict student performance will be examined [25]. Mayilvaganan and Kapalnadevi (2014) have published a paper on contrasted categorization methods for predicting pupils' academic success [12]. In the meantime, Gray et al.
al. (2014) looked at the accuracy of classification models in predicting tertiary education advancement [36].

Table 2: Decision Tree accuracy results

Method	**Attributes**	**Results**	**Authors**
Decision Tree	College assessments	76%	[22]
	Intelligence factors	65%	[36]
	CGPA	91%	[28]
	Interior valuation, CGPA, Extra-curricular events	66%	[12]
	Student Demographics, high school background	65%	[25]
	Intelligence aspects, Extra-curricular events, spoken skills	88%	[32]
	Internal valuation, Student Demographics, Extra-curricular activities	90%	[18]
	Student-Teachers Interaction	73%	[9]

3.2.2. Neural system

A neural system is a computer that makes judgments based on data. Another popular approach in educational data mining is neural networks. A neural network's ability to identify all possible interactions between predictor variables is one of its advantages.[36]. Even in composite nonlinear relationships among dependent and independent variables, a neural link could perform a comprehensive detection without any uncertainty [29]. As a result, one of the most effective prediction approaches is neural network technology.

As a result of the meta-analysis study, eight (8) articles using the Neural Network technique have been published. An Artificial Neural Network model for predicting student performance is provided in the publications. [38] [29]. Admission statistics [24], students' attitudes toward self-regulated learning, and academic success [19] Neural Network looks at various criteria. The rest are identical studies, but the researchers combined the two methodologies with the addition of the Decision Tree method.

Table 3: Neural Network accuracy results

Method	Attributes	Results	Authors
Neural Network	College assessments	81%	[38]
	Intelligence factors	69%	[36]
	External assessment	97%	[29]
	CGPA	75%	[16]
	GPA, Pupil Demographics, College Record, Allowance, Public interaction	71%	[9]
	Student Demographics, College Record	72%	[25]
	Exterior valuation, Student Demographics, College Record	74%	[24]
	College assessments, Exterior valuation	98%	[19]

3.2.3. Naive Bayes

A forecast can also be made using the Naive Bayes approach. There are a total of thirty (30) papers in this collection.

Naive Bayes algorithms are used in four studies to predict student achievement. We all want to achieve the same thing.

The aim of these four (4) papers is to recognize the most accurate technique for predicting student achievement. [9, 12, 25, 16] are examples of comparisons. According to their findings, Naive Bayes utilized all of the data's properties.

Then it examined each one to demonstrate the significance and independence of each attribute [9]. Table 4 displays the outcome.

Table 4: Naive Bayes accuracy results

Method	Attributes	Results	Authors
Naive Bayes	GPA, Pupil Demographics, College Record, Scholarship, Public interaction	76%	[9]
	Pupil Demographics, College Record	50%	[25]
	CGPA	75%	[16]
	College valuation, CGPA, Co-curricular events	73%	[12]

3.2.4 K-Nearest Neighbour

After examining three papers, K-Nearest Neighbour had the finest performance and accuracy, as indicated in Table 5. According to Bigdoli et al. (2003), the K-Nearest Neighbour approach uses smaller intervals to characterize students' performance as slow learners, regular learners, decent learners, or excessive learners. [23, 12]. When it comes to approximating the complete design of a learner's evolution through postsecondary schooling, the K-Nearest Neighbour technique has a high level of accuracy [36].

Table 5: K-nearest Neighbour accuracy results

Method	Attributes	Grades	Author
K-nearest Neighbour	Intelligence features	69%	[36]
	College valuation, CGPA, Co-curricular events	83%	[12]

	College assessment	82%	[23]

3.2.5 Support Vector Machine (SVM)

Support Vector Machine is a classification method that uses supervised learning. The Support Vector Machine was used to forecast student performance in three articles. Because it performs well with small datasets, Hamalainen et al. (2006) chose the Support Vector Machine as their forecast method [10]. According to Sembiring et al. (2011), the Support Vector Machine is more general and faster than previous approaches [35].

Meanwhile, Gray et al. (2014) discovered that when identifying students who are at risk of failing, the Support Vector Machine approach had the highest prediction accuracy.

The accuracy of prediction is seen in Table 6.

Table 6: SVM accuracy results.

Method	Attributes	Results	Author
SVM	Intelligence factors	83%	[35]
	College valuation, CGPA, Co-curricular events	80%	[12]
	College valuation, CGPA	80%	[10]

Prediction Accuracy Grouped by Algorithms:

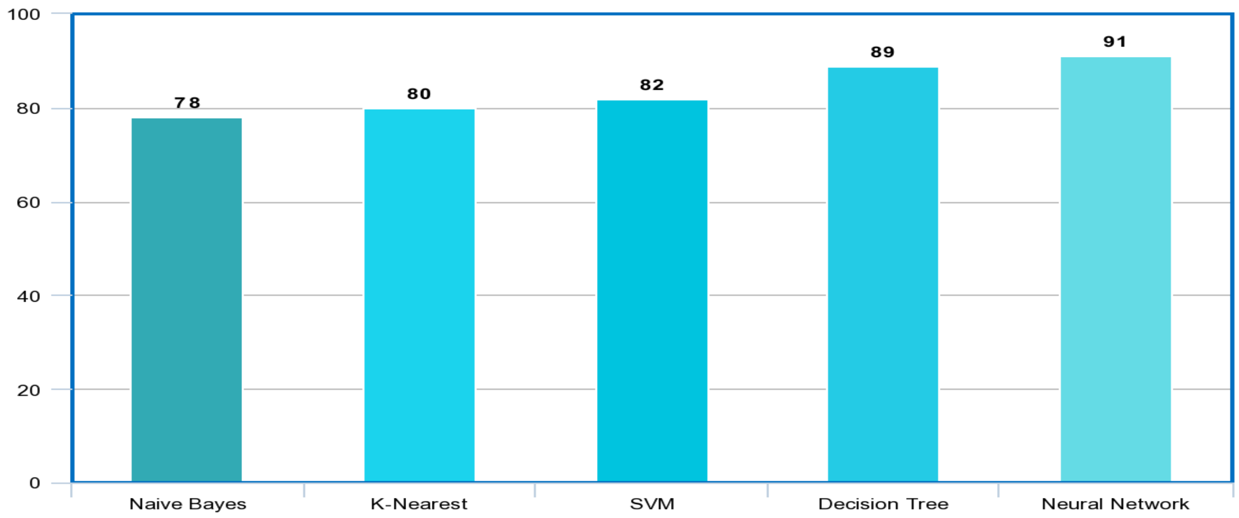

Fig. 2 Precision results assembled by algorithms from 2002-2015

4. Purposed Architecture

This system accepts input in the form of a textual dataset.

We know that we're doing data processing on the system, therefore we're employing data processing, modules.

- pre-processing
- feature extraction
- classification, all of which use our Random Forest (best performing) algorithm

So, First Input as a Textual dataset then pre-processed the dataset (pre-processing step is clean the dataset). After that, the system extracts the parameter in the dataset of student performance in the extraction section.

Then, in classification, where we utilize our Random Forest algorithm to classify and predict.
Output is to Detect Student Performance.

4.1 Architecture diagram

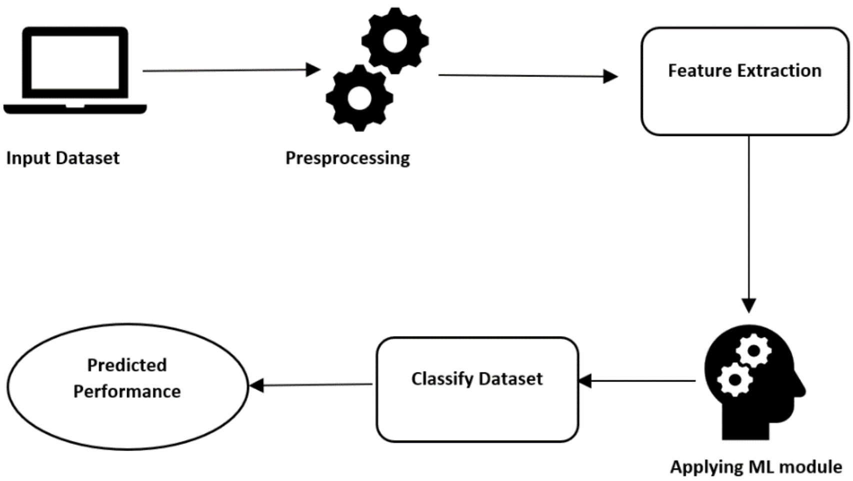

Fig. 3 System Architecture

4.2 Module wise Explanation

Figure 3 shows system architecture diagram.

Module 1: Login/ Registration

Input: Fill out the registration form and then login into our system.

Output: User will get ID and Password.

Module 2: Data Pre-processing

Input: Collect dataset of student performance and upload dataset into our system. Data pre-processing cleans the datasets by like removing raw data and finding missing values in the dataset.
Output: successfully clean dataset.

Module 3: Training

Input: In the training dataset, we will clean the dataset, extract features in the dataset and classify the dataset using machine learning techniques.
Output: Successfully trained dataset.

Module 4: Testing

Input: In testing, first create GUI and design pages. Import Random Forest Model and predict the output.
Output: Predict Student Performance.

5. Deliberation

This section will review current studies' findings on predicting student success. This metanalysis is based on the most accurate prediction methodologies and the essential aspects that could affect a student's performance. From 2002 to 2015, The accuracy of classification systems grouped by algorithms for forecasting student achievement is depicted in Fig. 2.
According to the graph in Fig. 2, the Neural Network has the highest prediction accuracy (98%), followed by the Decision Tree with a prediction accuracy of 98 percent. (91% of the time). The accuracy of the Support Vector Machine was then compared to that of the K-Nearest Neighbour algorithm (83%). Finally, Naive Bayes is a prediction method with a lower accuracy (76%).
The traits or features used throughout the prediction process determine the forecast's accuracy. Because of the influence of the main qualities, the Neural Network technique provided the best prediction accuracy.
Internal and external assessments are used to form these traits [5]. The accuracy is reduced by 1% when only one variable is used, external assessments [29]. Internal assessments were the third most commonly utilized variable, with an accuracy rate of (81%) [38]. It demonstrates that external assessment, or the grades received in the final test, plays a significant effect in forecasting student achievement. On the other hand, psychometric variables had the least impact on student

performance, with an accuracy of only (69%) [36]. Because psychometric criteria frequently involve qualitative data, forecasting with a Neural Network algorithm rather than quantitative data is difficult. The maximum error prediction of the Neural Network approach, on the other hand, is smaller. The most significant forecast inaccuracy is less than 10% [24]. Another advantage of neural networks is that nonlinear interactions are easily captured. It's also known as an adaptive system because of its ability to update past data as quickly as a human brain. As a result, the model always works outside of the knowledge base. Furthermore, a neural network's strength is its capacity to learn from a small data collection [11]. The Decision Tree approach is the second most accurate prediction method (91 percent) in terms of performance accuracy. [16]. The component that leads to the highest accuracy in anticipating students' achievement in the Decision Tree technique is CGPA. Two additional studies support this claim, both of which show that the results are better when CGPA is one of the primary criteria.

The performance prediction accuracy was around 90% [18, 13]. Therefore, it is possible to conclude that the Decision Tree is capable to work with both numerical and categorical data [12], to perform well in large datasets [28], and to be easy to understand and use. [39, 32] saw a relationship between variables. Aside from that, the least important element in predicting student achievement is parental involvement. perceived the link between factors Aside from that, the least important factor in predicting student success is performance is based on psychometric criteria [36], with a result of only 80% accuracy
(65%). It demonstrates that the Decision Tree is effective.
The next machine is the Support Vector Machine, which has a performance accuracy of approximately (83%). According to the findings, psychometric characteristics are the best attributes for using the Support Vector Machine approach to predict student success [10]. However, the result dropped to (73%) performance accuracy when extracurricular activities were added.
On the other hand, K-Nearest Neighbour displayed significant accuracy (83 percent) in predicting student achievement using three factors: internal evaluation, CGPA, and extracurricular activities.[12].
Decision Tree and Nave Bayes accuracy results are lower when compared to the other two approaches, Decision Tree and Nave Bayes. Compared to the K-Nearest Neighbour technique [12], Extracurricular activities were also utilized as a factor in another study. To improve forecast accuracy, they paired it with two other criteria attribute [13][32].

Finally, Naive Bayes is the approach with the lowest prediction accuracy (76%) [9]. The variables used are CGPA, student demographics, high school background, scholarship, and social network interaction. All of these characteristics were also Although both Neural Network and Decision Tree approaches used Naive Bayes, the results showed that Naive Bayes provided the best outcomes.
Compared to Neural Networks and Decision Trees, the accuracy of the Decision Tree is superior. This is because each of the attributes employed is significant.

6. Experiment Results

We have achieved highest accuracy of 94.62% with random forest algorithm we also tested our dataset with different algorithms such as Neural Network, SVM, K-nearest, Decision Tree.
The performance evaluation of the following model is done on 962 sample of data out of 962 sample of data 769 samples were used for the training purpose and 193 sample size was used for the testing purpose.
Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression
Random Forest has given us best accuracy and result.

6.1 Dataset

The dataset and understanding of the data are critical for determining the accuracy of the student performance system. The following system was developed to define the academic goal and objectives. The following dataset has 962 Samples of data it consists of different attributes such as Class, Gender, Semester, Nationality, Grade Id, Relation, Topic, Stage Id, Place of birth, Students Absent days, Parents Answering Survey, Parents Satisfaction with School.

Attributes	Values
Class	Primary School-0 Middle School-1 High School-2
Gender	Male-0 Female-2
Semester	First-1

	Second-2
Parents Answering Survey	Yes-0 No-1
Parents Satisfaction with School	Good-0 Bad-1
Students Absent Days	Under 7 - 0 Above 7 - 1
Relation	Father-0 Mom-1
Topic	IT-0 Maths-1 Arabic-2 Science-3 English-4 Quran-5 Spanish-6 History-7 French-8 Biology-9 Chemistry-10 Geography-11
Stage ID	Lower Level-0 Mid-Level-1 High School-2
Nationality	Kuwait-0 Lebanon-1 Egypt-2 Saudi Arabia-3 USA-4 Jordan-5 Venezuela-6 Iran-7

	Morocco-8 Syria-9 India-10 Libya-11
Place of birth	Kuwait-0 Lebanon-1 Egypt-2 Saudi Arabia-3 USA-4 Jordan-5 Venezuela-6 Iran-7 Morocco-8 Syria-9 India-10 Libya-11
Section ID	A-0 B-1 C-2
Grade ID	G-02=0 G-04=1 G-05=2 G-06=3 G-07=4 G-08=5 G-09=6 G-10=7 G-11=8 G-12=9

We used one hot encoding to categorize data into numerical order

6.2 Hardware & Software Requirements

Software Requirement:

Operating system: 64-bit Windows 10.

Coding Language: Python

Design constraints: Spyder

Hardware Requirement:

RAM: 8GB

Speed: 2.4 GHz

Hard Disk:500 GB

6.3 Mathematical Model

Let S be the Whole system which consists of:

S= {IP, Pro, OP}.

Were,

IP is the input of the system.

Pro is the procedure applied to the system to process the given input.

OP is the output of the system.

A. Input:

IP = {I}.

Were,

{I} is a set of the dataset provided as input.

B. Procedure:

Step1: pre-processing

Step2: Feature Extraction.

Step3: Classification (Using ML module)

Step5: Create Model

C. Output: show the result.

Table 7: Comparative analysis of proposed approach with existing systems

Year & Ref	Method	Accuracy
2014,2013,2015 [9, 12, 25, 16]	Naïve Bayes	78%
2003,2014 [23, 12]	K-Nearest	80%
2014 [36]	SVM	82%
2008 [22]	Decision Tree	89%
2008,2012 [24,19]	Neural Network	91%
Proposed	**Random Forest**	**94.62%**

7. Conclusion and Future Work

Predicting student performance is particularly beneficial when it comes to assisting educators and students in enhancing their learning and teaching processes. Prior research on employing various analytical tools to predict student performance was examined in this paper. Internal assessment and cumulative grade point average (CGPA) were being used as data sets for the majority of the research. In the field of educational data mining, the classification method is commonly utilized for prediction strategies.

In terms of classification strategies, researchers have shown that Neural networks and Decision trees are two of the most popular ways of predicting student success. In terms of classification strategies, researchers have shown that Neural networks and Decision Trees are two of the most popular ways of predicting student success. Finally, the meta-analysis on student performance prediction has inspired us to conduct additional research that can be applied in our situation.

It will help the educational institution maintain track of students' progress in a systematic way. We have taken into account the academic data of a large number of students; nevertheless, there are still a large number of students and a large amount of input data that might be used in the future. A lot of data can be taken into consideration for more accurate findings as to the demand for not only the student but also overall performance prediction grows. In the field of data mining, there is a lot of room for student performance prediction. In the Future, the Use of Different Classifier and Classify will help us to predict Student Performance in a more précised way.

References: -

[1] National higher education strategy plan, Malaysian Ministry of Education (2015).

[2] "An overview of applying academic analytics to anticipate and improve students' achievement: A proposed proactive intelligent intervention," by Usamah bin Mat and colleagues. 126-130 in 2013 IEEE 5th Conference on Engineering Education (ICEED).

[3] Predicting students' academic performance: comparing artificial neural networks, decision trees, and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007. Z. Ibrahim, D. Rusli

[4] "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601-618, November 2010, doi: 10.1109/TSMCC.2010.2053532.

[5] "Association Rule Generation for Student Performance Analysis Using Apriori Algorithm," by Angeline and D. Magdalene Delighta. (2013).

[6] Pragmatic cost estimation for online applications, M. W. Sukumar Letchuman and Mac Roper

[7] Rialette Pretorius, David Budgen, O. Pearl Brereton, Mark Turner, Mahmood Niazi, Stephen Linkman, Barbara Kitchenham

Volume 52, Issue 8, Systematic literature reviews in software engineering – A tertiary study (2010).

[8] Drop out feature of student data for academic performance using decision tree techniques, M. M. Quadri and N. Kalyankar, Global Journal of Computer Science and Technology 10 (2).

[9] Data mining approach for forecasting student performance, E. Osmanbegovic and M. Sulji' c, Economic Review 10 (1).

[10] Comparison of machine learning algorithms for intelligent teaching systems, W. Ham al ainen and M. Vinni, in Intelligent Tutoring Systems, Springer, 2006, pp. 525–534.

[11] Mining educational data to improve students' performance: a case study, M. M. A. Tair and A. M. El-Halees, International Journal of Information 2 (2).

[12] Comparison of classification algorithms for forecasting the performance of students' academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, pp. 113–118, M. Mayilvaganan and D. Kalpanadevi.

[13] Expert systems with applications 41 (14) (2014) 6400–6407. S. Natek and M. Zwilling, Student data mining solution–knowledge management system for higher education institutions, Expert systems with applications 41 (14) (2014) 6400–6407.

[14] T. M. Christian and M. Ayub, "Exploration of classification using nbtree for forecasting student performance," in IEEE International Conference on Data and Software Engineering (ICODSE), 2014, pp. 1–6.

[15] Predicting student academic performance, in: Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, IEEE, 2013, pp. 27–33. K. F. Li, D. Rusk, and F. Song, Predicting student academic performance, in: Complex, Intelligent, and Software Intensive Systems (CISIS),

[16] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, using optimal equal width binning and synthetic minority over-sampling technique to improve the accuracy of a student's final grade prediction model, Decision Analytics 2 (1) (2015) 1–25.

[17] S. Parack, Z. Zahid, and F. Merchant, "Application of data mining in educational databases for anticipating academic trends and patterns," in IEEE International Conference on Technology Enhanced Education (ICTEE), 2012, pp. 1–4.

[18] Elakia et al./(IJCSIT) International Journal of Computer Science and Information Technologies 5 (3) (2014) 4649–4652, Elakia et al./(IJCSIT) International Journal of Computer Science and Information Technologies 5 (3) (2014) 4649–4652.

[19] Appraising the significance of self-regulated learning in higher education using neural networks, International Journal of Engineering Research and Development Volume 1 (Issue 1) (2012) 09–15.

[20] Z. K. Papamitsiou, V. Terzis, A. A. Economides, Temporal learning analytics for computer-based testing, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 31–35.

[21] Mining student-generated textual data in moocs and quantifying their effects on student performance and learning outcomes, in: 2014 ASEE Annual Conference, Indianapolis, Indiana, 2014

[22] 'Data mining algorithms to classify students,' in: Educational Data Mining 2008, 2008, C. Romero, S. Ventura, P. G. Espejo, C. Hervas

[23] Predicting student performance: An application of data mining approaches with the educational web-based system lon-capa, by B. M. Bidgoli, D. Kashy, G. Kortemeyer, and W. Punch, in Proceedings of the ASEE/IEEE Frontiers in Education Conference, 2003

[24] Predicting students' academic performance using artificial neural networks: A case study of an engineering course, The Pacific Journal of Science and Technology 9 (1) (2008) 72–79. V. Oladokun, A. Adebanjo, and O. Charles-Owaba.

[25] Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013) 35–39. V. Ramesh, P. Parkavi, and K. Ramar.

[26] Clustering for better educational process mining, in: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, ACM, 2014, pp. 11–15. A. Bogar'n, C. Romero, R. Cerezo, M. Sanchez-Santill'an,

[27] Visualizing patterns of student involvement and performance in MOOCs, by C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, in Proceedings of the fourth international conference on learning analytics and knowledge, ACM, 2014, pp. 83–92.

[28] Data mining: Prediction for graduate student performance improvement using classification, in Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.

[29] P. M. Arsad, N. Buniyamin, and J.-l. A. Manan, A neural network students' performance prediction model (nnsppm), in IEEE International Conference on Smart Instrumentation, Measurement, and Applications (ICSIMA), 2013, pp. 1–5.

[30] Personality differences among entering male and female medical students, S. S. Meit, N. J. Borges, B. A. Cubic, and H. R. Seibel, Online Submission.

[31] Learning techniques of successful and failed university students., Online Submission 1 (1) (2010) 36–45. A. Simsek, J. Balaban

[32] "Mining Students' Data for Prediction Performance," 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 255-262, doi: 10.1109/ACCT.2014.105, T. Mishra, D. Kumar, and S. Gupta.

[33] Predicting students' final performance from participation in online discussion forums, Computers & Education 68 (2013) 458–472. C. Romero, M.-I. Lopez, J.-M. Luna, S. Ventura.

[34] Factorization models for forecasting student performance, N. Thai-Nghe, T. Horvath, and L. Schmidt-Thieme, Educational Data Mining 2011, 2010.

[35] Prediction of student academic performance using data mining approaches, in: International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, 2011, pp. 110–114. S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, E. Wani.

[36] An application of classification models to forecast student development in tertiary education, G. Gray, C. McGuinness, and P. Owende, Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 549–554.

[37] I. Hidayah, A. E. Permanasari, and N. Ratwastuti, "Student classification for academic performance prediction using neuro fuzzy in a traditional classroom," in IEEE International Conference on Information Technology and Electrical Engineering (ICITEE), 2013, pp. 221–225.

[38] Using neural networks to predict student achievement, T. Wang and A. Mitrovic, Computers in Education, 2002. Proceedings. IEEE International Conference on, pp. 969–973, 2002.

[39] Predicting ndum student's academic performance using data mining approaches, M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, Computer and Electrical Engineering, 2009. Second International Conference on, IEEE, 2009, Vol. 2, pp. 357–361.