RESEARCH ARTICLE

European Journal of Information Technologies and Computer Science
www.ej-compute.org

Model for Prediction of Student Grades using Data Mining Algorithms

Kdhaiya Sulaiman Al Shibli, Amal Sulaiman Sayed Al Abri, Linitha Sunny, Nandakishore Ishwar, Sherimon Puliprathu Cherian

ABSTRACT

There has been a rapid growth in the educational domain since education has become an important need. Data is collected in this domain which can be put to meaningful use to derive a lot of benefits to the students. Predicting student performance can help students and their teachers keep track of student progress. Mining Educational data helps to uncover invisible patterns, relationships, or trends in the unstructured data and helps in delivering logical and meaningful recommendations. Several kinds of research are being conducted across the world to analyze the data regarding student learning to identify the factors affecting performance and to provide support to students to help them improve. It is the objective of the proposed research to conduct a detailed study in the Sultanate of Oman regarding the existing toolsets, systems, and mode of data collection that are used currently in the Education sector for the prediction of Student Grades. Taking this as the baseline, later a model that will feature different prediction algorithms which are more accurate in predicting the grades of a student will be developed. The objective of this research is to understand the various predictive methods used to predict student performance and to propose a machine learning model to predict student grades.

Keywords: Classification techniques, data mining, educational data mining, student grade prediction.

Published Online: March 22, 2022

ISSN: 2736-5492

DOI :10.24018/ejcompute.2022.2.2.47

K. S. Al Shibli
Faculty of Computer Studies, Arab Open University, Muscat, Sultanate of Oman.
(e-mail: s160059@aou.edu.om)
A. S. S. Al Abri
Faculty of Computer Studies, Arab Open University, Muscat, Sultanate of Oman.
(e-mail: s160229@aou.edu.om)
L. Sunny
Department of Computer Science, Saintgits College of Engineering, Kerala, India.
(e-mail: linithasunny25@gmail.com)
N. Ishwar
Department of Computer Science, Saintgits College of Engineering, Kerala, India.
(e-mail: ishwarnandakishore@gmail.com)
S. P. Cherian*
Faculty of Computer Studies, Arab Open University, Muscat, Sultanate of Oman.
(e-mail: sherimon@aou.edu.om)

****Corresponding Author***

I. Introduction

Students are one of the valuable assets of a nation. There is a constant need to continuously improve the standard of education as well as the knowledge of students. Student marks or grades reflect their understanding or knowledge of subjects imparted through education. Analysis of this data using tools like data mining will help in gaining insight and predicting the marks or grades of a student and suggest guidelines for improvement. This will help in taking corrective measures well in advance for improving the student knowledge and standard of education. Educational institutions across the Sultanate collect vast amounts of data like student personal and education details, student marks, related educational documents, student performance during a period. Meaningful analysis and mining of this data can benefit the student community for their future improvement. In modern educational systems which impart education via different modes, it is becoming more and more complex to make accurate predictions of student performance. Tools that are being used currently to predict require a more sophisticated and intelligent approach to making more accurate predictions. So, we are proposing a website-based prediction interface that will allow students as well as faculty to view the predicted marks a student is likely to obtain in future examinations. By doing so, we provide a simple platform to assess the performance of each individual student, which accordingly will enhance the overall performance of the Institution as well as the country. The main purpose of this paper is to identify and understand the various predictive methods used to predict the student performance and to present a machine learning model to predict student grades.

II. Related Work

The study looked at research articles published between 2020 and 2021. The most popular repositories were examined to discover related resources, including Science Direct, Scopus, IEEE, and Google Scholar. The key terms "student pre-diction algorithms," "data mining algorithms in the educational area," "data min-ing for student performance," "student performance prediction," and "educational data mining" were used to search for articles. For reliable search results, the Boolean operator "AND" was employed.

European Journal of Information Technologies and Computer Science
www.ej-compute.org

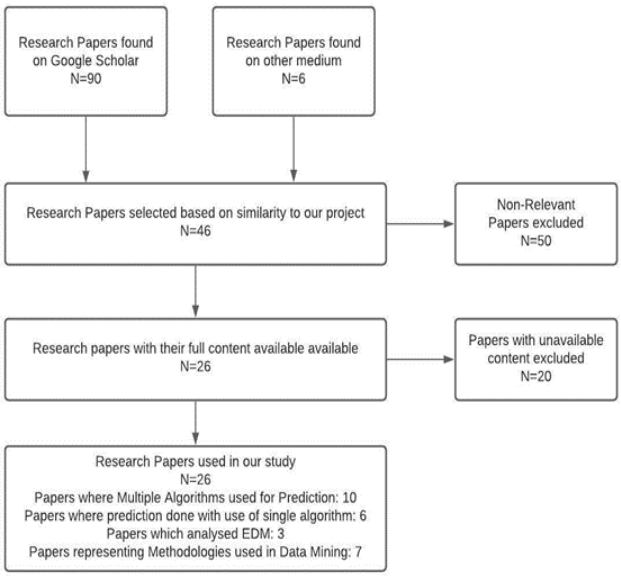

Fig. 1. Flowchart.

Initially, ninety research articles were in Google Scholar and six were collected from other sources during the search. There were 46 articles that were deemed to be like our topic, and 50 were excluded because they did not match our topic. Only 26 of the 46 articles had complete material, thus 20 of them were eliminated at this point. Multiple algorithms were employed for prediction in ten of the articles, a single algorithm in six, educational data mining was examined in three of the publications, and data mining approaches were discussed in seven of the papers. The flowchart of the search technique is shown in Fig. 1.

After reviewing the selected research papers, we found that in most cases, student issues affecting student performance are high school level, gender, parental education of students, location, financial background, student family status, intermediate teaching, students last semester marks, grade test, conference performance, assignment assignments, general skills, class and web work, Interest in a particular subject, learning style, time and family support in the course, type of admission, previous school marks, accommodation, parent qualifications, parental work. All the above factors fall into various categories, such as personal, family, academic, institutional, and social. Therefore, student performance can also be influenced by personal and academic life.

Reference [1] captured the test data of 6000 students and used a predictive data prediction method to study diploma student performance in school. The decision tree as a means of separation has been used and a predictive model has been developed. This model was able to show the highest relative accuracy of student performance. One such project involved the use of data mining techniques to design and build a Data Mining model that enables student performance predictions in educational institutions in Ethiopia. Student-related information is mined with a decision tree classifier to model using the C4.5 Decision tree algorithm [2]. This paper reviews EDM research and techniques from an educational perspective, the growth in the field of e-Learning, data sources for decision support, and improvement of teaching and learning practices [3]. Reference [4] studies the performance related to the education of the students. The data focus relating to students namely those achieving low grades and high grades.

Reference [5] used and gave a brief overview of the algorithms of SVM, Naive Bayes, and Decision Tree. The Naive Bayes is said to be a separator that simply uses a method of classification for machine learning opportunities. Also, they described SVM as an indirect machine, a supervised learning model. In the Decision Tree, it kept the division of data into simple units. In this paper, we found out that the mid-term marks and final marks played a huge role. Also, their analysis showed that class test- 2 and individual presentation marks played a substantial role in getting exceptional grades for the students. In their findings, SVM had more accuracy (89.12) than Decision Tree (85.81) and Naive (87.8187), However, the Support Vector Machine took more time than the other two algorithms which were 0.88s and 0.86s. Overall Naive Bayes is the best algorithm they worked on having a good trade-off between time and accuracy.

Reference [6] took into consideration Artificial Neural Networks (ANN), decision tree, SMO, Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbor, Random Forest, Random Tree, LADTree, Linear Regression, J48, REPTree algorithms. Here the authors went through many research papers and found that CGPA and the student's internal marks in the subjects are important indicators of predicting results. Also, they strongly believed that the extraction of education data would play an important role in predicting student performance, predicting future school dropouts, predicting institutionalization and admission to the new academic year is very helpful for teachers, administrators, and education policymakers.

Reference [7] highlighted a Java-based data mining tool called WEKA, developed at the University of Waikato, New Zealand. The algorithms of the various categories used in WEKA are NaiveBayes, Decision Tree, RandomTree, JRip, and ZeroR, and many other algorithms pre-installed in it, making it easy for the user. They also found that academic data mining researchers used CGPA and internal performance marks to predict course performance. However, in their initial results, they found that school, as well as study time, also affects students' final grades. Class algorithms such as OneR, JRip, and Decision Tree have more than 80 percent accuracy in predicting student outcomes, which are equally effective.

In this paper, the G-SA algorithm, which is a new hybrid of GA and SA algorithms, is presented to predict students' academic performance [8]. The G-SA algorithm has been compared to five well-known classification algorithms in data mining, including Naive Bayes (NB), Multilayer-Perceptron (MP), Decision Tree, J48, LR, and SVM. Furthermore, it has been compared to the four following meta-heuristic methods, Basic ABC algorithm, GBC algorithm. Basic SA algorithm, SA-GA algorithm. In summary, a new method called G-SA is presented for the prediction of students' academic performance during educational courses, which can be used to prevent possible failures of students. The G-SA algorithm uses the advantages of both simulated annealing and genetic algorithms by combining them. The combination of the two

algorithms also balanced the power of exploration and exploitation in the proposed algorithm, which had not only helped to speed convergence but has also been able to get rid of local optimum. Experimental results from the implementation of the G-SA algorithm show that the proposed algorithm improves accuracy performance from 1.09 to 24.39 percent compared to other metaheuristic comparison methods and from 0.29 to 6.57 percent compared to well-known conventional classification methods.

Reference [9] provided us with information on new data mining tools such as RapidMiner, KNIME, Orange, and many more that are constantly evolving. RapidMiner is a package for data mining analysis and modeling. KNIME is a data purification and analysis package that will assist in the data processing. Orange is a package for data recognition and analysis. SPSS is a mathematical package used for statistical testing, regression frameworks, integration, and asset analysis. KEEL is used for analysis and has algorithms for classification and regression. This paper presented a study of the tools used in EDM and a review of current EDM trends in which related work strategies in the field are compared. Reference [10] In this study, they used the Orange data mining platform as an open source software for data mining and machine learning. Data mining strategies include Linear Regression, Rules Association, Naive Bayes, Decision Tree, and Random Forest. Separation and regression methods were used to predict student performance. While the organizational rules process has been used to find the most common items in student records to understand the reasons for their failure. In addition, it identified weaknesses and failures of students and assessed information that helps to improve the education system. Furthermore, it has tried to find reasons for repeated student failures in a particular subject.

Reference [11] use machine learning variants such as Passive-Aggressive Classifier (PAC), Linear Discriminant Analysis (LDA), RadiusNeighborsClassifier (RNC), Support Vector Machine (SVM), Extra Tree (ET). The best accuracy found between these different processes was 94.86 percent from SVM. The second highest accuracy obtained was 93.21 percent in the case of LDA. They did not consider achieving the high accuracy of classifiers as a good measure of the partition algorithm. The purpose of the proposed project was to create a more efficient framework that significantly improved student performance. Reference [12] used WEKA (Waikato Environment for Information Analysis) as a testing tool for data classification. Working on this paper was very difficult with tree size, entropy, and other factors. This project also created a website such as how we plan and use WEKA software. Test results show that the true positive rating for the POOR, AVERAGE, GOOD, VERY GOOD, EXCELLENT class is 79.1, 80.3, 87.8, 97.2, and 100 percent respectively, and J48 is the best data classification algorithm used in this study with an accuracy of 85.6471 percent.

Reference [13] gave a brief explanation and understanding of prediction, classification, clustering, and association. This paper explains DM (data mining), its applications, and techniques used in the education sector. The prime intention of the paper was to showcase the true value of data mining in the education sector which it did nicely. It also discusses some tools developed for applying these techniques to education data that can also be used to detect the dropout reasons which we have taken inspiration to add to our system. Reference [14] quoted Educational Data Mining as "Education data mining is a leading data mining application which is a data mine set to answer questions in terms of the educational environment that illuminates the learning process.". Here they discussed all the methods available in Educational Data Mining and that Mining Data and Educational Data Mining are not very different in terms of features. In this paper, they have introduced various methods that can be used to search for data on education and the problems they may face in data mining. Their focus was on measuring student performance, learning behavior, and finding out to improve personal results. Such strategies will help to improve the overall outcome of the educational organization and at a higher level as well. Reference [15] have shown how Data Mining is divided into several categories such as data cleansing, data integration, etc. It clarified that the most widely used algorithms in Educational Data Mining are Decision Tree, Naive Bayes Classifier, and Neural Network. Among them, the Naive Bayes had the best accuracy, so it was selected in this study.

Reference [16] focuses on the use of planning programs to analyze students' performance in finding hidden patterns, which aid officials in making important educational decisions. Data mining procedures were applied to retrieve hidden information from student data, and comparisons were made between three dividers, Nave Bayes, Decision Tree and Random Forest. They also described that the data acquisition method involves four stages: data collection, pre-analysis, mining process (classification), and explanation of results. Test results demonstrated that Random Forest surpassed other categories with an accuracy of 71.3 percent, followed by Random Forest by 69.8 percent, with the last division being the Naïve Bayes with 59.4 percent. This research has aided us in predicting students' performance to improve their future achievement by depending on previous results. Also, they advised educational institutions to provide an education advisor to students who have failed to improve their academic performance. Reference [17] discussed about Fast Correlation-Based Filter (FCBF) which was selected as the feature selection algorithm. Their study spoke of the fact that Selecting Feature helps EDM to create a predictable student prediction model. They also examined the performance of the FCBF which led to the results saying that the FCBF performed satisfactorily with a multi-faceted student database.

In [18], 15 classifier algorithms were used to predict CGPA for students. These 15 were Decision Tree-based dividers (such as Random Tree, Decision stump, J48, Hoeffding Tree, LMT, REP Tree and Random Forest), rule-based dividers such as (PART, Jrip, and ZeroR), KNN, Naive Bayes, logistics, SMO, and simple logistics. But here they had a difference as they used all the partitions with the default setting except KNN. At KNN, they used automatic search to find the best value for k (i.e., the best number of neighbors). The median accuracy of all 15 categories was 71

percent. Hoeffding Tree, Naive Bayes, SMO, LMT, Random Forest, IBK and Simple Logistic, were seven categories that were above the 71 percent line. Two dividers, Hoeffding Tree and Naive Bayes had a good result of 91 percent accuracy. SMO had an accuracy of 87 percent, Random Forest had an accuracy of 86 percent, LMT had an accuracy of 84 percent, and Simple Logistic had an accuracy of 84 percent. In view of this, we can say that the top performance of tree-based separators was attained by Hoeffding Tree by 91 percent and then by Random Forest and LMT. IBK (i.e., KNN) obtained 76 percent accuracy, and Random Tree, J48, and REP Tree tree dividers obtained 96 percent, 65 percent, and 63 percent accurately, sequentially. In addition, classifiers based on the PART and Jrip law found 68 percent and 67 percent straight, respectively. The most unpleasant side effects, 48 percent, and 49 percent were obtained by Logistic and Decision Stump.

Reference [19] aims to review the most common methods of classification to predict student performance. The decision tree method has proven to be an effective way to determine student performance when compared to Naive Bayes, nearby neighbor, vector support systems, and artificial networks, among the most used models of predicting student performance, owing to its simplicity of use and ability to detect small or large data structures and foresee prices. They are precise. Reference [20] focused more on how to understand data that are generated dynamically in complex form. In conclusion, they stated that the solution to such a problem is the usage of data mining techniques in a profound manner. The advice was also given to implement data mining algorithms for the decision-making and analysis of the students in every aspect.

Reference [21] emphasized on algorithm-based parameters such as Accuracy, Probability Threshold, Execution Time, Precision, Recall, and other parameters. They conclude by saying that the entire data mining method and algorithm results vary depending on the data and the variable attribute used for prediction. However, when we use decision tree algorithms such as JRip, ADTree, Ridor, regression logistic, and neural network approach, depending on the specific needs these algorithms provide incredibly accurate prediction results and help to improve the education system.

Reference [22] compared the classification method and predict student performance. This has provided an opportunity for middle school students to improve their grades and to know the in-depth analysis of their shortcomings in terms of their performance and what they need to do to improve. In addition, this paper demonstrated the relative performance of C4.5, ID3, Random tree calculation, Naïve Bayesian classifier calculation, and to have the correct precision of the layout calculation, and choice tree calculation to analyze student performance.

Reference [23] offered a data-mining method to assess the legitimacy of a hypothesis by performing a hypothetical analysis to see if CGPA final and grade level of students in their final year can be determined using their GPA from the first three years of their studies. The system and year of entrance were utilized to incorporate KNIME workflow estimations utilizing six data mining approaches in a performance comparison analysis of one of six distinct algorithms. R2 values of 0.955 and 0.957 were found using both line-based and pure regression order models, indicating a high level of accuracy of 89.15 percent. This demonstrates that the achievements of engineering students' graduation in their fifth and final year of study conducted in Nigeria may be anticipated based on their achievement in the first three Math.

III. Materials and Methods

A. Dataset

For the study we have considered a dataset that has considered several features that might affect their future grades. The input features considered are:

```
data.columns
```

```
Index(['school', 'sex', 'age', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu',
       'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime',
       'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery',
       'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc',
       'Walc', 'health', 'absences', 'G1', 'G2', 'G3'],
      dtype='object')
```

Fig. 2. Dataset attributes.

These features are used to predict the target features G3(grade). The dataset was divided into training set and test set in the ratio 70:30.

B. Proposed System

The proposed system uses various concepts of Machine Learning that would help in predicting the grade of a student based on the history of the student's previous performance so that corrective measures can be taken well in advance for the performance improvement of the student.

This will help in taking corrective measures well in advance for improving the student knowledge and standard of education. Educational institutions across the World collect vast amounts of data like student personal and education details, student marks, related educational documents, student performance during a period. Meaningful analysis and mining of this data can benefit the student community for their future improvement. In modern educational systems which impart education via different modes, it is becoming more and more complex to make accurate predictions of student performance. Tools that are being used currently to predict require to be more sophisticated and have an intelligent approach to making more accurate predictions.

Based on the literature review we conducted, the algorithms listed below proved to be an effective and provided better accuracy than others in determining student performance:

- Linear Regression
- Naive Bayes
- Random Forest Classifier
- Support Vector Machine

1) Linear Regression

It is the most basic type of regression. By fitting a linear equation to the data, linear regression seeks to model the relationship between two variables. The goal of linear

regression is to discover the mathematical relationship between two variables. If the result is a straight line, it is a linear model; if it is a curved line, it is a nonlinear model. A straight line represents the link between dependent variables, and there is only one independent variable.

$Y = \alpha + B\,X$

Model 'Y' is a linear function of 'X'.

The value of 'Y' increases or decreases in linear manner according to which the value of 'X' also changes.

2) Naive Bayes

The Naive Bayes method is a supervised learning algorithm for addressing classification issues that is based on the Bayes theorem. It is mostly utilized in text classification tasks that require a large training dataset. The Naive Bayes Classifier is a simple and effective classification method that aids in the development of fast machine learning models capable of making accurate predictions. It's a probabilistic classifier, which means it makes predictions based on an object's probability.

Spam filtration, sentiment analysis, and article classification are all common uses of the Naive Bayes Algorithm.Bayes' theorem, often known as Bayes' Rule or Bayes' law, is a mathematical formula for calculating the probability of a hypothesis given previous information. It is conditional probability that determines this.

The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (1)$$

where, P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B. P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true. P(A) is Prior Probability: Probability of hypothesis before observing the evidence. P(B) is Marginal Probability: Probability of Evidence.

3) Random Forest Classifier

A random forest is a supervised machine learning system that uses decision tree algorithms to build it. There are three parts to a decision tree: decision nodes, leaf nodes, and a root node. A decision tree method separates a training dataset into branches, each of which is further divided into branches. This pattern repeats until a leaf node is reached. There is no way to separate the leaf node any farther.

The attributes utilized to forecast the outcome are represented by the nodes in the decision tree. The leaves are connected to the decision nodes.

4) Support Vector Machine

A Support Vector Machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors.

IV. RESULTS

The correlation between different features is represented in Fig. 3. Table I shows the prediction accuracies of various algorithms used during the study.

From the below results obtained it can be concluded that Linear Regression gives the maximum prediction accuracy.

TABLE I: ACCURACY OF PREDICTIONS

Algorithm	R2 Score
Linear Regression	0.8153645884693685
Naive Bayes	0.4006302599803363
Random Forest Classifier	0.7091359231565973
Support Vector Machine	0.5357081537321841

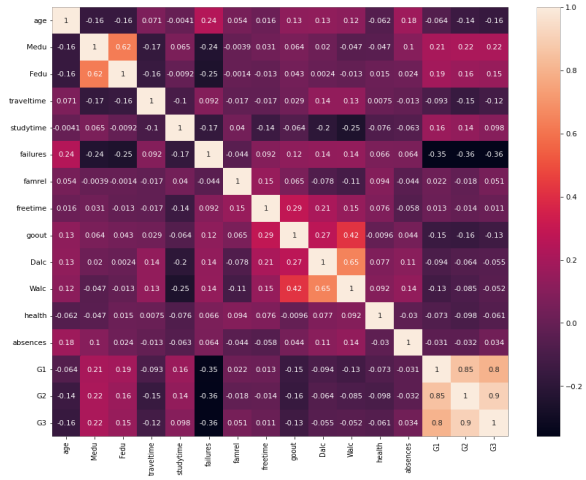

Fig. 3. Correlation matrix.

Linear Regression R2 Score:

```
[ ]  model = LinearRegression()
     model.fit(X_train, y_train)

     LinearRegression()

[ ]  print(f"Model R2: {model.score(X_test, y_test)}")

     Model R2: 0.8153645884693685
```

Fig. 4. R2 score – Linear Regression.

Naive Bayes R2 Score:

```
[ ]  from sklearn.naive_bayes import GaussianNB
     gnb=GaussianNB()
     gnb.fit(X_train,y_train)
     y_pred=gnb.predict(X_test)
     print(y_pred)

     [16 12  7 14  9 12 12 14 14 14  7  7 14 13 14 16 19 16  6 14  6  7 13  8
      16 14  6  9 14 16 14  7 16 14  7  7 16 16  7 12 14 16 16  6 18 14  9  7
       8 14  7 14 14 14 13  7 14  8 16  7 16  6 14 16 16  8  9 11 14  7  7  6
      11 14 16  9  6  6  8 12  0  7  6 13 14  8  6 16 14 16  8  6 19  7  7  6
      14  0  6  7 12 16  7 14  8  0  7 12  8  7  6  7  0 13 16 14  7  7  0]

[ ]  print(f"Model R2: {model.score(X_test, y_pred)}")

     Model R2: 0.4006302599803363
```

Fig. 5. R2 score – Naïve Bayes.

Random Forest Classifier R2 Score:

```
[ ]  from sklearn.ensemble import RandomForestClassifier
     rf = RandomForestClassifier()
     rf.fit(X_train,y_train)
     R_pred=rf.predict(X_test)
     print(R_pred)

     [14 11  0 13  9 10 12 13 11 12  0  8 10 12 10 15 15 15 10 11 11 10 13 10
      15 10 14  7 14 15 10 10 15 10 10 13 12 15  0 12 13 15 15 10 15 13  0 13
       8 15 10 15 11 11 11  0 10 10 14 11 10  6 11 14 13  6 10 10 11  0 10 12
       9 11 14  9 10  0 10 13  0 11 10 13 15  8 10 14 13 15 10 11 14  9  8 11
      13  0  6 10 10 14 10 15 10  0  8 13 10  8 10  8 18 11 15 11  8 10  0]

[ ]  print(f"Model R2: {model.score(X_test,R_pred)}")

     Model R2: 0.7091359231565973
```

Fig. 6. R2 score – Random Forest.

Support Vector Machine R2 Score:

```
[ ]  from sklearn import svm
     classifier = svm.SVC(kernel='linear',gamma = 'auto', C=2)
     classifier.fit(X_train,y_train)
     y_predict = classifier.predict(X_test)

[ ]  print(f"Model R2: {model.score(X_test, y_predict)}")

     Model R2: 0.5357081537321841
```

Fig. 7. R2 score – Support Vector Machine

V. Conclusion

Assessment of students' performance is a continuous process that forms an integral part of the education process. Student resources are the fuel for any future progress of a nation. A high-performing student community leaves a good impact in all the sectors where they join or are engaged in. They are the future technologists, scientists, and innovators who influence the future growth of the nation. Improving student performance holds the key to better educational practices and effective educational decisions. Predictive Learning and Data Analytic involve steps by which future prediction is easily possible based on the current and historically relevant educational data of the student community. This article gives a brief detail on how students' mark predictions were done using different methodologies and a machine learning model for predicting student grades.

From our research, we gained a lot of knowledge regarding Data Mining and its various aspects of it including Educational Data Mining. We will use this to its benefit during the next stages of our project. The future work includes creating a user-friendly website that can be accessed by the educators of their institutions where when given a data set as input the predicted grades for the next academic examinations will be displayed for each Data Mining Algorithm. Finally, we conclude that studying the current work in terms of predicting student performance has prompted us to do some research work in our field of study. It will greatly aid in the improvement of our educational system by allowing us to assess a student's usual performance.

Funding

The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program BFP/URG/ICT/20/216.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References

[1] Al Mazidi A, Abusham E. Study of general education diploma students' performance and prediction in Sultanate of Oman, based on data mining approaches. *Int. J. Eng. Bus. Manag*., 2018; 10: 1847979018807020.

[2] Tegegne AK, Alemu TA. Educational data mining for students' academic performance analysis in selected Ethiopian universities. *Inf. Impact J. Inf. Knowl. Manag*. 2018; 9(2).

[3] Educational Data Mining: A review of evaluation process in the e-learning - ScienceDirect. [Internet] [cited 2021 Oct 30] Available from: https://www.sciencedirect.com/science/article/abs/pii/S0736585317306639

[4] Analyzing undergraduate students' performance using educational data mining - ScienceDirect. [Internet] [cited 2021 Oct 30] Available from: https://www.sciencedirect.com/science/article/abs/pii/S0360131517301124

[5] An Analysis of Students' Academic Record Using Data Mining Techniques and Identification of Key Factors to Aid Students' Performance | *GUB Journal of Science and Engineering*. [Internet] [cited 2021 Oct 30] Available from: https://www.banglajol.info/index.php/GUBJSE/article/view/47900

[6] Kumar M, Singh AJ, Handa D. Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques. *Int. J. Educ. Manag. Eng*. 2017; 7(6): 40–49.

[7] Walia N, Kumar M, Nayar N, Mehta G. Student's Academic Performance Prediction in Academic using Data Mining Techniques. *Proceedings of the International Conference on Innovative Computing & Communications (ICICC). 2020.*

[8] Rohani Y, Torabi Z, Kianian S. A Novel Hybrid Genetic Algorithm to Predict Students' Academic Performance. *J. Electr. Comput. Eng. Innov. JECEI*. 2020; 8(2): 219–232.

[9] Mahajan G, Saini B. Educational Data Mining: A state-of-the-art survey on tools and techniques used in EDM. *Int. J. Comput. Appl*. 2020; 12(1): 7.

[10] Al-Fairouz EI, Al-Hagery MA. Students Performance: From Detection of Failures and Anomaly Cases to the Solutions-Based Mining Algorithms. *Int. J. Eng. Res. Technol*. 2020; 13(10): 2895,

[11] Singh R, Pal S. Application of Machine Learning Algorithms to Predict Students Performance. *Int. J. Adv. Comput. Res*. 2020; 29: 7249–7261.

[12] Hoque M, Azad A, Tuhin AH, Salehin Z. University Students Result Analysis and Prediction System by Decision Tree Algorithm. *Adv. Sci. Technol. Eng. Syst. J*. 2020; 5: 115–122.

[13] Mushtaq S, Singh SN. Analysis of Data Mining Methodologies & Applications in Education Sector. *Proceedings of the International Conference on Innovative Computing & Communications (ICICC).* 2020.

[14] Patel RA, Patel B, Garg D, Kaka B. Educational Data Mining: An Application of Data Mining To Estimate Students' Performance. 2020.

[15] Prediction of student learning outcomes using the Naive Bayesian Algorithm (Case Study of Tama Jagakarsa University) - IOPscience. [Internet] [cited 2021 Oct 30] Available from: https://iopscience.iop.org/article/10.1088/1757-899X/823/1/012056/meta

[16] Al-Fairouz E, Al-Hagery M. The Most Efficient Classifiers for the Students' Academic Dataset. *Int. J. Adv. Comput. Sci. Appl*. 2020; 11: 501–506.

[17] Role of FCBF feature selection in educational data mining | *Mehran University Research Journal of Engineering & Technology*. [Internet] [cited 2021 Oct 30] Available from: https://search.informit.org/doi/abs/10.3316/INFORMIT.459135063399758

[18] Alangari N, Alturki R. Predicting Students Final GPA using 15 Classification Algorithms. 2020.

[19] Lynn ND, Emanuel AWR. Using Data Mining Techniques to Predict Students' Performance. a Review. *IOP Conf. Ser.: Mater. Sci. Eng.* 2021; 1096(1): 012083.

[20] Kaur H, Bathla EG. Student Performance Prediction Using Educational Data Mining Techniques. *Int. J. Future Revolut. Comput. Sci. Commun. Eng*. 2018; 4(12).

[21] Ashraf A, Anwer S, Khan MG. A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques. *Am. Acad. Sci. Res. J. Eng. Technol. Sci*. 2018; 44(1).

[22] Kumar A, Eldhose KK, Sridharan R, Panicker VV. Students' Academic Performance Prediction using Regression: A Case Study. *International Conference on System, Computation, Automation and Networking (ICSCAN).* 2020: 1–6.

[23] The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *ScienceDirect.* [Internet] [cited 2021 Oct 30] Available from: https://www.sciencedirect.com/science/article/pii/S240584401836924X