Predicting Students' Academic Performance Using Naïve Bayes

Abdullah Baz, Fatima Alshareef, Ebtihal Alshareef, Hosam Alhakami, Tahani Alsubait
aobaz01@uqu.edu.sa, fbshareef@uqu.edu.sa, alshareefeb1@gmail.com, hhhakam@uqu.edu.sa, tmsubait@uqu.edu.sa
College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia

Abstract
Nowadays, universities and many educational institutions have a critical responsibility towards society, shaping multiple social factors. Therefore, it is very important to predict the educational output of these institutions at early stages. It is very challenging to predict students' academic performance because of the huge bulks of data stored in the environments of educational databases. Students' performance can be predicted with the help of various available techniques. Data Mining is the most prevalent family of techniques to predict students' performance and is extensively used in the educational sector, referred to as Educational Data Mining. In this paper, a dataset is collected from Umm Al-Qura University database. This dataset consists of 138 records of students who graduated from College of Computer and Information Systems in the year 2019, associated with 13 attributes including student ID, gender, eight courses' grades, GPA of both first and second semester in the first's year and the final GPA. The classification algorithm called Naïve Bayes is employed on the dataset by using the WEKA tool. Results achieved show that Naïve Bayes can be used for predicting students' academic performance at early stages in the first year with an accuracy of 72.46%.

Key words:
Educational Data Mining; Performance Prediction; Naïve Bayes

1. Introduction

Educational Data Mining (EDM) is an emerging field of applying data mining algorithms in the educational environment to develop learning and teaching processes. Educational data has been growing rapidly [1], hence the main aim of EDM algorithms is to analyze the huge data to solve educational issues.

The process of EDM is composed of tasks to be performed in sequence. Data collection refers to the collection of data that are used by educational institutions such as personal, enrollment and academic information. Data pre-processing refers to the pre-processing of data to resolve incomplete and/or inconsistent data issues and transfer data into useful information. EDM algorithms apply educational data mining algorithms for analyzing data and gaining useful insights. These algorithms contain statistics and visualizations to understand the results of useful analysis. Data interpretation refers to the interpretation of analyzed data to get the required conclusion. It helps in data identification and for future predictions. On the basis of analysis and interpretation of data for evaluating the process implementations comes the modification of the EDM process [2]. Different techniques and algorithms support the process of EDM [3]. The most used data mining techniques are classification, clustering, regression, association and sequential patterns.

Nowadays, universities and many educational institutions have a critical responsibility that affect all the society factors. Therefore, it is very important to predict the educational output of these institutions at early stages. The aim of predicting performance of the student is developing the educational processes.

Old is not always gold. For the aforesaid, sometimes we need to change old methods, syllabuses and techniques [4]. There are many influences that had an important role to conduct this main goal. By figuring out the students' weaknesses and strengths, they can improve their performance and avoid the academic failure. For professors, they may change their methods to be suitable for the students' abilities. As well as, for the curriculum committees and admission committees, they can set the appropriate criteria for the admission and improve the syllabuses.

The aim of our research is to predict students' academic performance at Umm Al-Qura University by using Naive Bayes method, one of the most known data mining classification algorithms. This classifier helps to predict the final GPA of students at early stages based on courses' grades in the first year. This paper is organized as follows. In Section 2, we present a collection of studies that investigate analyzing students' performance using data mining algorithms. Section 3 describes the proposed methodology. Section 4 illustrates our experimental results, while Section 5 presents some thoughts for future work.

2. Related work

In this section, we include many studies which were published between 2019 and 2020. These studies have a lot of contributions in predicting students' academic performance using several educational data mining techniques. Various criteria were taken into consideration, namely objective, algorithm, tool, dataset, attributes and prediction accuracy. In order to clarify the weaknesses and strengths to improve educational process. Table 1 summarizes these studies.

Manuscript received April 5, 2020
Manuscript revised April 20, 2020

Table 1: A comparison of related work

Year	Objective	Algorithm	Tool	Attribute	Dataset	Highest accuracy	Reference
2019	Predicting student's performance	PNN, Random Forest, Decision Tree, Naïve Bayes, Tree Ensemble, Logistic Regression	WEKA	GPA for the first three academic years and the final CGPA	1,841 students from 2002-2014 across 7 engineering departments in a Nigerian University	89.15% for Logistic Regression	[5]
		Naive Bayes, Random Forest, JRip, REPTree, OneR, J48, Simple Logistic and ZeroR	WEKA	Thirty-three attributes including academic grades, demographic attributes, social attributes and school related attributes	649 student's data from two secondary school of Portuguese	76% for J48, REP-Tree and OneR	[6]
		K Nearest Neighbor, Random Forest	Python	CGPA, Quantitative Aptitude, Coding Languages Known, English Speaking Skills, Number of projects and Internships done	306 students' data in higher education of Kalinga Institute of Industrial Technology for the final year academic batches	93.54% for Random Forest	[7]
		Random Forest, Tree Ensemble, Decision Tree, Naive Bayes, Logistic Regression, and Resilient back propagation	KNIME and Orange platforms	student's entry age, the aggregate WAEC score, JAMB score, university based CUSAS score, first-year grade classification while the actual CGPA was considered for the regression analysis	1445 student records from2005 to 2009 in their first year at Covenant University in Nigeria	50.23% for Logistic Regression in KNIME, 51.9% for Neural Network in Orange	[8]
		Random Forest, Decision Tree, Tree Ensemble, Gradient Boosted Tree, k-NN and Support Vector Machine	KNIME	Student name, ID, Gender, course' grade, Semester GPA, State of the students based on semester GPA, and final CGPA	398 business student's data from the Marketing department of a renowned university in Bangladesh from 2013 to2016	94.1% for Random Forest	[9]
		K-NN, Naive Bayes, Decision Tree and Logistics Regression	-	name, ID, gender, CGPA, and all the courses enrolled by the students including the course' grade	631 students from Faculty of Computer and Mathematical Sciencesat Universiti Teknologi MARA Cawangan Kelantan and Universiti Teknologi MARA Cawangan Negeri Sembilanat from 2013 to 2016	89.26% for Naïve Bayes	[10]
		J48, NNge and MLP	-	33 attributes (student grades, demographic, social and school related features).	1044 student from two schools in Alentejo region Portugal for 2005-2006 academic year	95.78% for J48	[11]
		ID3 and J48	WEKA	13 attributes (Father's Income, Mother's Education, Mother Working Status, Student's Study Hours, Tuition, Social Network Usage)	500 students from various departments of University of Ghana	62.67%fort he hybrid classification algorithm	[12]
		Neural Network-MLP, SVM, K-NN classifiers, Decision Tree, Naïve Bayes, Random Forest and Multi-class Classifier	WEKA and Rapid Miner	11 attributes (ID, Raised-hands, Visited Resources, Announcements View, Discussion, Parent	1100 student from Saudi University database	100% for the Random Forest	[13]

				Answering Survey, Etc.)			
	Predicting students' dropout	J48, Random Tree, REP-Tree, OneR, ZeroR, and JRip	WEKA	admission method, major, education status, term of enrollment, grade point average of university, province of high school, and grade point average of high school	4,238 records from Faculty of Science, Prince of Songkla University from 2013 to 2017	77.30% for JRip	[14]
		Neural Network (NN), Decision Tree, Support Vector Machine and kNN	MATLAB	13 attributes including institutional, academic, demographic, psychological and financial factors	481 students at a case study university of males as well as females	83.7% for NN	[15]
2020	Predicting students' academic performance	J48 and Naive Bayes	WEKA	Exams Marks, GPA, School, Sex, Age, Nationality, and City	38671 students' data of both male and female from Umm Al-Qura University, Saudi Arabia in the last 5 years	84.38% for J48	[16]
		JRip, NNge, OneR, Prism, Ridor, J48, Simple Cart, AD Tree, Random Tree and REP Tree	WEKA	15 attributes	1,268 students of three schools in Colombia in the 2018-2019 academic year	98.5% for ADTree	[17]
		Decision tree J48, Naïve Bayes and K-Nearest Neighbor	WEKA	age, sex, organization involved in school, extracurricular activities, pocket money, duration of study, duration of social media, duration of playing online games, information on attendance, illness, permission, semester grades one and two	253 students of SMA Negeri3 Ambon	99.6047% for decision tree algorithm J48	[18]
		PPP, L-SVM, R-SVM, GP, DT, RF, NN, ADB, NB and clustering techniques	-	open date of an assignment, date of first view of the assignment, date of assignment submission and due date of the assignment	242 students from the University of Tartu in Estonia	96% for PPP	[19]

In 2019, 11 EDM papers were published, nine of them used to predict students' performance and two to predict students' dropout. To predict students' performance, Adekitan and Salau [5] applied six data mining algorithms. The dataset consists of the GPA data for the first three academic years and the final CGPA of 1,841 students from 2002 to 2014 across 7 engineering departments in a Nigerian University. The used tool was KNIME analytics platform to analyse the students' performance dataset. The result of their study showed that Logistic Regression had the highest prediction accuracy of 89.15%. In another study, Salal et al. [6] implemented data mining classification algorithms including Naive Bayes, Random Forest, JRip, REPTree, OneR, Decision Tree (J48), Simple Logistic and ZeroR for predicting students' academic performance. They collected 649 student's data with 33 attributes including academic grades, demographic attributes, social attributes and school related attributes from two secondary school of Portuguese then they analyzed it using WEKA tool. The result showed that Decision Tree (J48), REPTree and OneR had the prediction accuracy of more than 76%. Decision Tree (J48) had accuracy of 76.2712%, REPTree and OneR had the same accuracy of 76.7334%. The study [7] by Agarwal et al. analyzed 306 students' data in higher education of 306 students of Kalinga Institute of Industrial Technology for the final year academic batches using Python. They selected several attributes such as CGPA, Quantitative Aptitude, Coding Languages Known, English Speaking Skills, Number of projects and Internships done. By using two classification algorithms K Nearest Neighbor and Random Forest, the result of their study showed that Random Forest had the highest prediction accuracy of 93.54%. Adekitan and NomaOsaghae [8] used data mining algorithms including Random Forest, Tree Ensemble, Decision Tree, Naive Bayes, Logistic Regression, and Resilient back propagation in KNIME and Orange platforms. They analyzed student's data in their first year at Covenant University in Nigeria with several features such as student's entry age, the aggregate WAEC score, JAMB score, university based CUSAS score, first-year grade classification while the actual CGPA was considered for the regression analysis. The result of their study showed that Logistic Regression in KNIME platform and Neural Network in Orange platform had the prediction accuracy of 50.23% and 51.9% respectively.

In addition, Rifat et al. [9] used six classification algorithms of data mining including Random Forest, Decision Tree, Tree Ensemble, Gradient Boosted Tree, K-Nearest Neighbors (K-NN) and Support Vector Machine for predicting the students' performance. They collected 398 university student transcripts from 2013 to 2016 with the attributes including Student name, Student ID, Gender, All the courses including course' grade, Semester GPA, State of the students based on semester GPA, and final CGPA, then they analyzed it using KNIME, the Konstanz tools. The result of their study showed that Random Forest had the highest prediction accuracy of 94.1%. Yaacob et al. in their study [10] applied data mining algorithms such as K-NN, Naive Bayes, Decision Tree and Logistics Regression to predict student's performance. They collected data of 631 students from Faculty of Computer and Mathematical Sciences who have completed their academic degrees from 2013 to 2016 with attributes such as student name, ID, gender, CGPA, and all the courses enrolled by the students including course grades. The result of their study showed that Naive Bayes had the highest prediction accuracy 89.26%. The study [11] by Imran et al. have also tried to predict performance of 1044 students from two schools in Alentejo region Portugal for the 2005-2006 academic year with 33 attributes including student grades, demographic, social and school related features. They used three classification algorithms and the result showed that the J48 algorithm achieved highest accuracy 95.78%. Kumar et al. in their study [12] they predict the academic performance of 500 students from various departments of University of Ghana with 13 attributes. They used a hybrid classification algorithm of ID3 and J48 and WEKA tool that give accuracy of 62.67%. Another study [13] by Sultana et al. analyzed 1100 student from Saudi University database with 11 attributes. They used eight classification algorithms, WEKA tool and Rapid Miner tools. As a result, the Random Forest gives the highest accuracy of 100%.

On the other hand, to predict students' dropout, Pattanaphanchai et al. [14] proposed a model using six classifiers including J48, RandomTree, REPTree, OneR, ZeroR, and JRip. The dataset was collected from Faculty of Science, Prince of Songkla University of 4,238 records. They selected 7 attributes including admission method, major, education status, term of enrollment, grade point average of university, province of high school, and GPA. WEKA tool was used for machine learning algorithms. The result of their study showed that JRip had the best prediction accuracy of 77.30%. Also, Al-Sudani and Palaniappan in their study [15] analyzed 481 students' data at a case study that included males as well as females with 13 attributes including institutional, academic, demographic, psychological and financial factors. They used MATLAB and data mining algorithms including Neural Network (NN), Decision Tree, Support Vector Machine and k-Nearest Neighbors (K-NN). The result of their study showed that NN had the highest accuracy of 83.7%.

In 2020, we found four studies discussing the utilization of data mining techniques for solving problems in educational environments. Alhakami et al. [16] used Naive Bayes and J48 algorithms for predicting students' academic performance and help in advising students using WEKA tool. They collected 38671 students' data of male and female from Umm Al-Qura University in the last 5 years, with many attributes including Sex, Age, Nationality, City, Exams Marks, School and final grade. The result of their study showed that J48 had the highest prediction accuracy of 84.38%. Another study [17] by Viloria et al. used 10 classification algorithms and WEKA tool to predict the academic status at the end of the first semester for 1,268 students from the Preparatory Program of three schools in Colombia in the 2018-2019 academic year with 15 attributes. The result showed that ADTree gives highest accuracy of 98.5%. In their study, Pattiasina et al. [18] predict the performance of high school students with a dataset of 253 students with fourteen attributes. They used four classification algorithms and the WEKA tool. The result showed that decision tree algorithm J48 has the highest accuracy that is 99.6047%. Moreover, Hooshyar et al. [19] proposed a novel algorithm called PPP for automatic assessment of students' performance through procrastination behaviors by using their assignment submission data. The dataset was collected at 2019 of 242 students from the University of Tartu in Estonia with four attributes from the logs of the courses. They also use clustering techniques and eight classification algorithms. The result showed that the novel algorithm gives 96% accuracy.

Table 2: Dataset Description

Attribute	Description	Possible Values
ID	Student' ID	Unique number
Gender	Student's Gender	Male, Female
4800221-6	English Language Grade	A+, A, B+, B, C+, C, D+, D, F
4800223-2	Computer Skills1 Grade	A+, A, B+, B, C+, C, D+, D, F
4800256-4	General Physics (1) Grade	A+, A, B+, B, C+, C, D+, D, F
4800745-4	Introduction to Mathematics (1) Grade	A+, A, B+, B, C+, C, D+, D, F
4800011-4	Introduction to Maths (2) Grade	A+, A, B+, B, C+, C, D+, D, F
4800015-3	Computer Programming Skills Grade	A+, A, B+, B, C+, C, D+, D, F
4800242-3	Learning and Study Skills Grade	A+, A, B+, B, C+, C, D+, D, F
4800322-4	Technical English Language Grade	A+, A, B+, B, C+, C, D+, D, F
1' S GPA	First Semester GPA	Excellent, Very Good, Good, Pass
2' S GPA	Second Semester GPA	Excellent, Very Good, Good, Pass
Final GPA	Final GPA	Excellent, Very Good, Good, Pass

In total, 15 papers have been reviewed, with 45 contributing authors and 30 data mining identified techniques. The common attributes that were considered in these studies are academic grades, demographic attributes, social attributes and school related attributes. In addition, the highest accuracy obtained by Random Forest with 100% and Decision Tree with 99.60%.

3. Methodology

In this section, we present the stages of applying the data mining method for predicting students' academic performance.

3.1 Data Collection

In this study, we gathered records of bachelor students in Computer Science who graduated from College of Computer and Information Systems at Umm Al-Qura University, Saudi Arabia in the year 2019. The collected data was organized in Microsoft Excel sheet. In total, we have 138 students' records of both male and female. Each record has attributes namely student ID, gender, graduation year and semester, major, 8 courses taken by the student including the course' code, name and grade, GPA (Grade Point Average) of both first and second semester in the first's year and final GPA.

3.2 Data Preparation

During this step, we focused on preparing data to be suitable for the data mining process. Data cleaning is a step to delete missing values, those students who did not have 100% complete information were extracted from the dataset and we ignored the unneeded attributes that do not affect the predicting process.
According to Table 2, the selected attributes are 'ID' represents the student' ID, and 'Gender' means the students student's Gender. The attributes '4800221-6' means English Language , '4800223-2' Computer Skills1, '4800256-4' General Physics (1), '4800140-4' Introduction to Mathematics (1), '4800745-4' Introduction to Maths (2), '4800015-3' Computer Programming Skills, '4800242-3' Learning and Study Skills, '4800322-4' Technical English Language, all these attributes contains the student's grades. '1' S GPA' and '2' S GPA' represents the GPA of first level and second in the first year. 'Final GPA' means the final GPA of students.
Next, we changed course grades from numerical format (values from 0 to 100) to nominal format (values as A+, A, B+, B, C+, C, D+, D and F) as well as student's GPA was transformed into an Excellent, Very Good, Good and Pass using functions in MS Excel. This is because Naive Bayes algorithm gives better results with the categorical data rather than numerical data.

3.3 Tool Used

WEKA is a well-known tool used for machine learning and data mining that was developed by Waikato University in New Zealand. It contains a collection of tools for classification, regression, association rules, clustering, data pre-processing and visualization [20]. WEKA tool is widely used in academic and industrial environments.

3.3 Naive Bayes

Naive Bayes classification uses Bayes Theorem that calculates the probability of an event based on conditions that relates to the event [21]. It assumes that the classification attributes are independent considering the value of the class. Also, it works well with the categorical data [22]. Naive Bayes is simple and tends to learn quickly. In addition, it does not require large amount of training data. Bayes' theorem is a mathematical formula stated as follows:

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

Where:
P(A): the probability of A.
P(B): the probability the probability of B.
P(B|A): the probability of event B based on A condition.
P(A|B): the probability of event A based on B condition.

Naive Bayes algorithm helps to classify the students' academic performance based on their grades in eight courses of the first year. It works with nominal attributes (i.e., A, B, C, etc.). The students are classified into four different classes. These classes are taken into consideration as well as the attributes are used to calculate the probability that the students' performance will get event of classes that are Excellent, Very Good, Good and Pass.

4. Experimental Results

We loaded our dataset in WEKA then we obtained useful knowledge about some attributes before applying our data mining method by using the visualizing technique. For example, we showed that 51.44% of the graduated students had an excellent and 36.95% had a very good final GPA, while the performance of 10.14% of students was good and the rest had pass GPA as shown in the Figure 1.

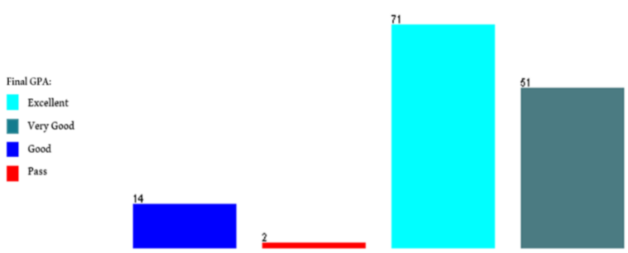

Fig. 1 Final GPA of Students.

Moreover, we analyzed the relationship between the final GPA of students and their grades in the first year's courses. For example, Figure 2 demonstrates the students' grade in Computer skills course. As it shows, the majority of the students who received an A+ or A was graduated with an "Excellent" final GPA. Also, most of the students earned a B+ or B in this course they graduated with a "Very Good" GPA. In addition, student who received C+ or C had a "good" GPA, while a low ratio of students who earned a D+ or D graduated with a "Pass" final GPA.

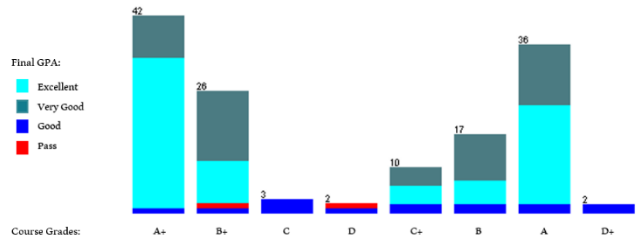

Fig. 2 Relationship Between Students Grades in Computer Skills Course and Their Final GPA.

Furthermore, according to Figure 3 and Figure 4, the final GPA of students reflect the GPA of both first and second semesters. For instance, students who graduated with an excellent GPA achieved an excellent GPA in the first and second semester. Also, students who had Very Good GPA kept their performance at the same level as they graduate. In addition, students whose performance was Good or Pass in both semesters in the first year graduated with the same performance. Therefore, we can realize that the final GPA can be predicted by the students' academic performance in the first year.

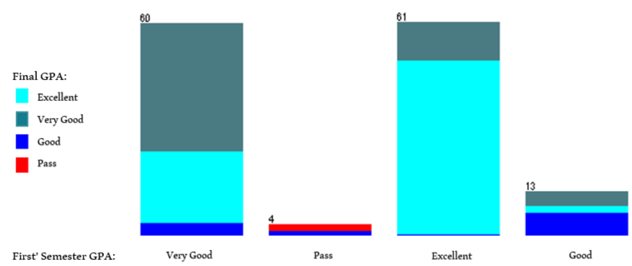

Fig. 3 Relationship Between First's Semester GPA and Final GPA.

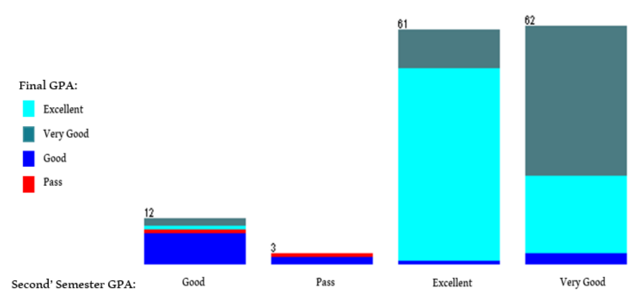

Fig. 4 Relationship Between Second's Semester GPA and Final GPA.

We formatted the data to Attribute-Relation File Format (ARFF), because WEKA prefers to load data in this format. Then, we obtained the dataset that will be analyzed. Our dataset was divided into training and testing data files randomly. The sample dataset in Table 3 clearly shows 13 attributes as discussed above in order to be used for the classification that involved four classes of students' performances in this study.

After analyzing, Fig. 5 illustrates the results in Naïve Bayes classifier, the percentage of correctly classified instances is 72.46%. On the other hand, the percentage of incorrectly classified instances is 27.53%. In addition, we obtain the percentage of accuracy by four classes: Excellent, Very Good, Good and Pass as shown in Table III. These percentages are based on the following accuracy measurement factors:

- True Positive (TP) Rate: The number of instances that are truly classified for each class.
- False Positive (FP) Rate: The number of instances that are falsely classified for each class.
- Precision: The number of truly classified instances divided by the total number of classified instances.
- Recall: The number of classified instances divided by the total number of instances for each class.
- F-Measure: The average between precision and recall.

The challenge of analyzing by Naive Bayes algorithm lies in inability to predict Pass class. To improve the accuracy of this class, more attributes and different classification algorithms can be used. From this point, many researchers will motivate to find more accurate results

Table 3: Detailed Accuracy by Class

Class	TP Rate	FP Rate	Precision	Recall	F-Measure
Excellent	0.778	0.152	0.848	0.778	0.812
Very Good	0.704	0.190	0.704	0.704	0.704
Good	0.500	0.063	0.429	0.500	0.462
Pass	?	0.029	0.000	?	?

To wrap up, this paper applied Naive Bayes classification method to analyze students' academic performance at Umm Al-Qura University with 13 attributes with an emphasis on the final GPA of the student. The students' performance classified into four classes Excellent, Very Good, Good and Pass. Figure 6 illustrates the results in Naive Bayes classifier, the percentage of correctly classified instances is 72.46%. On the other hand, the percentage of incorrectly classified instances is 27.53%. As a result, the knowledge based on precision accuracy showed that Naive Bayes is able to predict Excellent class with 84.8%, 70.4% for the Very Good class and 42.9% for the Good class. However, the model was unable to predict the Pass class.

```
=== Summary ===

Correctly Classified Instances          50               72.4638 %
Incorrectly Classified Instances        19               27.5362 %
Kappa statistic                          0.5321
Mean absolute error                      0.1462
Root mean squared error                  0.3545
Relative absolute error                 48.9017 %
Root relative squared error             93.6349 %
Total Number of Instances               69
```

Fig. 5 Percentage of Classified Instances using Naïve Bayes.

5. Conclusion and Future Work

Education is an essential part of any society. EDM methods allow the extraction of information from raw data at a high level, offer interesting possibilities for the educational domain. Specifically, various studies have used EDM algorithms to enhance the quality of education and avoid academic failures.

In this paper, we have addressed the issue of predicting students' academic performance at Umm Al-Qura University, based on the final GPA. The utilized dataset consists of 138 students with 13 attributes. Classification is done in order to predict students in different class categories like Excellent, Very Good, Good and Pass. The classifier used was Naive Byes for classifying students. The analysis of results showed that Naive Byes can be used for predicting students' academic performance at early stages in the first' year. Accuracy achieved using WEKA and Naive Bayes algorithm is 72.46%.

In future work, we aim to enhance our study by using different classification methods with several attributes in order to increase classification accuracy and predict the Pass class. Various research direction can be made in the field of educational data mining to improve the classification accuracy using different classification algorithms. This helps academic instructors and academic institutions to make appropriate decisions and take appropriate actions to improve the performance of students.

References

[1] H.-C. Hung, I.-F. Liu, C.-T. Liang, and Y.-S. Su, "Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education," Symmetry, vol. 12, no. 2, p. 213, 2020.

[2] A. Shende, M. S. Thakare, S. Byagar, and M. A. N. Joshi, "The review of different data mining tools, techniques and algorithms for the data mining in education," Our Heritage, vol. 68, no. 27, pp. 290–296, 2020.

[3] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing automl in educational data mining for prediction tasks," Applied Sciences, vol. 10, no. 1, p. 90, 2020.

[4] R. Katarya et al., "A review: Predicting the performance of students using machine learning classification techniques," in 2019 Third International conference on I-SMAC (IoT in

Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 36–41, IEEE, 2019.

[5] A. I. Adekitan and O. Salau, “The impact of engineering students’ performance in the first three years on their graduation result using educational data mining,” Heliyon, vol. 5, no. 2, p. e01250, 2019.

[6] Y. Salal, S. Abdullaev, and M. Kumar, “Educational data mining: Student performance prediction in academic,” IJ of Engineering and Advanced Tech, vol. 8, no. 4C, pp. 54–59, 2019.

[7] K. Agarwal, E. Maheshwari, C. Roy, M. Pandey, and S. S. Rautray, “Analyzing student performance in engineering placement using data mining,” in Proceedings of International Conference on Computational Intelligence and Data Engineering, pp. 171–181, Springer, 2019.

[8] A. I. Adekitan and E. Noma-Osaghae, “Data mining approach to predicting the performance of first year student in a university using the admission requirements,” Education and Information Technologies, vol. 24, no. 2, pp. 1527–1543, 2019.

[9] M. R. I. Rifat, A. Al Imran, and A. Badrudduza, “Educational performance analytics of undergraduate business students,” International Journal of Modern Education and Computer Science, vol. 11, no. 7, p. 44, 2019.

[10] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacob, and N. M. Sobri, “Supervised data mining approach for predicting student performance,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, no. 3, pp. 1584–1592, 2019.

[11] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, “Student academic performance prediction using supervised learning techniques,” International Journal of Emerging Technologies in Learning (iJET), vol. 14, no. 14, pp. 92–104, 2019.

[12] A. Kumar, R. Selvam, and V.Palanisamy, “Prediction of student performance using hybrid classification,” International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 4, pp. 6566 –6570, 2019.

[13] J. Sultana, M. U. Rani, and M. Farquad, “Student’s performance prediction using deep learning and data mining methods,” International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 1, pp. 1018 – 1021, 2019.

[14] J. Pattanaphanchai, K. Leelertpanyakul, and N. Theppalak, “The investigation of student dropout prediction model in thai higher education using educational data mining: A case study of faculty of science, prince of songkla uni-versity,” Journal of University of Babylon for Pure and Applied Sciences, vol. 27, no. 1, pp. 356–367, 2019.

[15] S. Al-Sudani and R. Palaniappan, “Predicting students’ final degree classification using an extended profile,” Education and Information Technologies, vol. 24, no. 4, pp. 2357–2369, 2019.

[16] H. Alhakami, T. Alsubait, and A. Aljarallah, “Data mining for student advising,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 3, 2020.

[17] A. Viloria, J. G. Guliany, W. N. Nu´nez, H. H. Palma, and L. N. ˜ Nu´nez, “Data mining applied in school dropout prediction,” in ˜ Journal of Physics: Conference Series, vol. 1432, p. 012092, IOP Publishing, 2020.

[18] T. Pattiasina and D. Rosiyadi, “Comparison of data mining classification algorithm for predicting the performance of high school students,” Jurnal Techno Nusa Mandiri, vol. 17, no. 1, pp. 22–30, 2020.

[19] D. Hooshyar, M. Pedaste, and Y. Yang, “Mining educational data to predict students’ performance through procrastination behavior,” Entropy, vol. 22, no. 1, p. 12, 2020.

[20] E. G. Kulkarni and R. B. Kulkarni, “Weka powerful tool in data mining,” International Journal of Computer Applications, vol. 975, p. 8887, 2016.

[21] S. Menaka and G. Kesavaraj, “Predicting student performance using data mining techniques: A survey of the last 5 years,” International Journal of Advanced Scientific Research and Management, vol. 4, no. 1, pp. 98–102, 2019.

[22] M. M. Saritas and A. Yasar, “Performance analysis of ann and bayes classification algorithm for data classification,” International Journal of Intelligent Systems and Applications in Engineering, vol. 7, no. 2, pp. 88–91, 2019.

Abdullah Baz received the B.Sc. degree in electrical and computer engineering from UQU, in 2002, the M.Sc. degree in electrical and computer engineering from KAU, in 2007, and the M.Sc. degree in communication and signal processing and the Ph.D. degree in computer system design from Newcastle University, in 2009 and 2014, respectively. He was a Vice-Dean, and then the Dean of the Deanship of Scientific Research with UQU, from 2014 to 2020. He is currently an Assistant Professor with the Computer Engineering Department, a Vice-Dean of DFMEA, the General Director of the Decision Support Center, and the Consultant of the University Vice Chancellor with UQU. His research interests include VLSI design, EDA/CAD tools, coding and modulation schemes, image and vision computing, computer system and architecture, and digital signal processing. Since 2015, he has been served as a Review Committee Member of the IEEE International Symposium on Circuits and Systems (ISCAS) and a member of the Technical Committee of the IEEE VLSI Systems and Applications. In 2017, IEEE has elevated him to the grade of IEEE Senior Member. He served as a Reviewer in a number of journals, including the IEEE Internet of Things, the IET Computer Vision, the Artificial Intelligence Review, and the IET Circuits, Devices and Systems.

Fatima Alshareef received the B.S. degree in Computer Science from Dammam University in 2014. Currently, she is a Master of Science candidate at Umm Al-Qura University, College of Computer and Information Systems, Computer Science Department. She now works as Teaching Assistant at Umm Al-Qura University.

Ebtihal Alshareef received the B.S. degree in Computer Science from Umm Al-Qura University in 2015. Currently, she is a Master of Science candidate at Umm Al-Qura University, College of Computer and Information Systems, Computer Science Department. From 2015-present, she works at King Saud bin Abdulaziz for Health Science University.

Hosam Alhakami received his B.Sc. degree in Computer Science from King Abdulaziz University, Saudi Arabia in 2004. From 2004 to 2007, he worked in software development industry, where he implemented several systems and solutions for a national academic institution. Following that, he started his postgraduate studies in UK, where he received his MSc degree in Internet Software Systems from Birmingham University, Birmingham, UK in 2009. Then he successfully acquired his PhD in Software Engineering from De Montfort University in 2015. His research interests include algorithms, semantic web and optimization techniques. He focuses on enhancing real-world matching systems using machine learning and data analytics in a context of supporting decision-making.

Tahani Alsubait is a faculty member of College of Computer and Information Systems. She earned her PhD in AI and instruction from the University of Manchester. She hold a Bachelor's in Computer Science from King Saud University and a Master's from King Abdulaziz University. Her research interests include knowledge representation and reasoning, data analytics and HCI.