 https://doi.org/10.24191/jeesr.v23i1.008

Student Performance Classification: Data, Features and Classifiers

Muhammad Hareez Mohd Zaki, Mohd Azri Abdul Aziz*, Suhana Sulaiman and Najidah Hambali

***Abstract*—Recognising student's performances are one of the factors that contribute to the Educational Data Mining (EDM) which make it important in determining the students-related factors that affect their performance. Thus, a review on the student performance classification based on its data, features and classifiers has been made. The purpose of the study is to review on classification of student performance, data and features that are used in analysing student performance and also the classifiers that are used in classifying the performances of the student. By reviewing these factors, the insight of the data and features used during classification of student performance can be obtained. There are also many different types of classifiers that were used in student performance classification. A few classifiers had been reviewed to gain understanding in their performance towards the data classification of students' performance.**

***Index Terms*—Student performance classification, data, features, classifiers.**

I. Introduction

THERE had been a growing interest in student performance classification in the past few years [1]–[15]. Classification is a set of approaches that learn from data samples to create a model that can infer a special attribute called the class label given other explanatory variables called predictor variables [11]. The model is also known as classifier [11]. Commonly, a few classifiers such as Naïve Bayesian [3], [4], [9]–[11], Decision Tree [4]–[6], [9]–[11], Random Forest [3], Neural Network [4], [6], [10], [11], Support Vector Machine [7], [8], [11] etc were used on classifying the performance of students. The increase in the number of research in student performance classification causes the classification to become a technique that is sought-after by the educational institutes to use it as an analysis to study the future output when using the students' previous and current data [9]. Thus, having a classification technique is important in researching the data of students' performance.

This manuscript is submitted on 3rd November 2021 and accepted on 23rd June 2022. M.A.A.Aziz, S.Sulaiman, and N.Hambali are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor (azriaziz@uitm.edu.my).

Muhammad Hareez Mohd Zaki was with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA. (email: 2020480662@isiswa.uitm.edu.my).

*Corresponding author
Email address: azriaziz@uitm.edu.my

Prior to the students' performance classification, collecting students' performance data is essential in the data-gathering stage. The collected data from any educational settings such as Learning Management System (LMS), traditional classroom-based learning sessions, etc., are used to understand the students' academic performance and the process is also known as Educational Data Mining (EDM) [12]. The goal of EDM is to create techniques for analysing various data gathered from the education area [13]. It was stated in [10] and [14] that successful and effective decisions can be made by using this technique due its significance in decision making. Consequently, it will improve the performance of the students in education.

During the pre-processing stage, the features of the data are used to analyse the performance of the students. The use of features in each previous research is varied depending on the research's needs as there is no specific criteria in evaluating students' performance [4]. According to [15], the researchers normally used a few main features to evaluate the students' performance consisted of academic features, demographical features, and psychometric features. However, there were previous research on students' performance [2], [16] that produced different results despite having some similar features. A conclusion from research [4] stated that the study on students' performance should be done in its educational environment so that the improvement can be made to the education system to check the students' performance regularly. It is important for research to be carried out using its current environment or area so that the type of data and features used in the study can be improved further in investigating the students' performance.

In this paper, a review has been made on students' performance classification done by previous researchers. Several perspectives were discussed on the students' performance consisted of data and features that were used during the analysis and also the classifiers that were applied to classify the information on the topic. This research makes contributions to researchers and practitioners in students' performance classification by presenting a clear review on the data, features and classifiers that were explored when analysing the topic.

The rest of this paper is organised as follows. Section II presents on the data used in previous research. Section III elaborates the features used when analysing students' performance. Section IV will be on the classifier used when classifying the students' performance. Finally, the conclusion on the paper is presented in section V.

II. DATA

Countless number of research had been done in studying the student performances. Every research has its own specific quantity of data and type of observation data that are used for data mining. A different amount of data used will have different effect on the performance of the classifiers during data classification. However, some classification algorithms can perform well when using large amount of data while some of them are better when using moderate or small amount of data [17]–[19]. Furthermore, the type of observation data used also need to be considered. Most of the research used similar observation data such as students. The combination of different type of observation data in research such as normal students and students with disabilities will affect the overall result of the study as there is clearly a bias on normal students to perform better than disabled students. Thus, it is important to have similar observation data in research to avoid the error stated.

Table 1 shows a few research that studied the students' performance. A few examples are listed in each column of the table. The table has five columns that consisted of reference, observation data, attribute name, attribute type and values.

Normally, in a dataset, there are two variables used consist of independent and dependent variables. Independent variables are the input features of the research that are not influence by other variables such as age, gender and study hours [25]. On the other hand, dependent variable also known as responding variable is the result or outcome of a research when processing the independent variables. Thus, dependent variable will be greatly influenced by the changes in the independent variables.

There are also some researchers that stated the type of feature that are used in their studies as in reference [2]. Even though the other researchers did not describe the type of features they used, the attributes can be determined by analysing and observing the features [2] [16]. There are two main type of features that are usually used by the researchers which are categorical and numeric variables. This can be further broken down into two classes each. For categorical, there are nominal and ordinal variable while for numeric variable, there are discrete and continuous variable. In reference [2], all features that were used in the study were nominal variable. Nominal variable is a variable with no ranking sequence. Ordinal is the vice versa of nominal an as example size of an object that can be categorised as small, medium, and large. Discrete variable is numeric variable with only integer number. Continuous variable is a numeric variable that is measured in continuous scale such as weight. In the end of this section, many attribute type will be discovered.

Based on the Table 1, the quantity of observation data used by the previous research were different. All research except research [16] and [2] used the observation data in the range of 200 to 500 students. The research, [16] and [2] exceeded the amount of observation data performed by the other studies and both used large quantity of data which are 10330 students. From the conducted research, there are no specific numbers or amount of observation data that will ensure the accuracy of the classification model. However, according to [26] dataset size will affect the performance of the classification model. It was also stated that having a large dataset produced accurate classification performance and the consequences of having small datasets was overfitting problem [26]. Hence, one of the factors that affect the classification model's performance is the dataset which includes both independent and dependent variables.

TABLE I
PREVIOUS STUDIES ON STUDENTS' PERFORMANCE

Ref.	Data	Attribute Name	Attribute Type	Values
[16]	10330 Students with 14 attributes	Gender	Nominal	Male, Female
		Admission Score	Numerical	0.00-35.98
		Current Semester	Nominal	1-10
[20]	500 Students with 10 attributes	Parental Status	Nominal	Both, Mother only, Father only
		Father Occupation	Nominal	Cooley, Farmer, Weaver, Private, Government, Business, Not Applicable
		Mother Occupation	Nominal	House Wife, Cooley, Farmer, Weaver, Private, Government, Not Applicable
[21]	200 Students with 17 attributes	SSC Medium	Nominal	English, Native
		SSC Grade	Nominal	SSC Grade {Excellent, Good, Average, Poor}
		INTER Medium	Nominal	Medium of study, Other than English consider as Native. {English, Native}
[2]	10067 Students with 14 attributes	Age	Nominal	29 distinct values
		Admission Exam Score	Numerical	0.00-6.00
		Student Class	Nominal	Weak, Strong
[22]	234 Students with 13 attributes	Marital Status	Nominal	Married, Single
		Religion	Nominal	Muslim, Christian
		Nationality	Nominal	Nigerian, Foreigner
[23]	480 Students with 17 attributes	Section ID	Nominal	-
		Grade Level	Nominal	-
		Student Absent Days	Nominal	Above 7 days, Below 7 days
[24]	239 Students with 6 attributes	First Semester Grade	Nominal	A, B, C, D, E, F
		Class Test Score	Nominal	Poor , Average, Good
		Assignment Completed	Nominal	Yes, No

III. FEATURES

Commonly, in students' performance classification, many researchers had used Educational Data Mining to study and gain more information regarding the topic [2], [16], [20]–[24], [27]–[32], etc. Studies on students' performance are still progressing until the present day as they are important to universities or any educational sector to cope with the growing development of the students throughout their studies.

There are no specific criteria or features that can study students' performance [4]. Each of the students has varieties of personalities and backgrounds with a different history that may influence the student's performance in the future. For that reason, many different possible features can be considered during the studies. These were also proven when previous research showed their features to be analysed to study the students' performance. Suppose the accuracy of the classification model is high. In that case, the features used in the experiment can be referenced because of the great results of the features on the model.

A. Demographical Features

In addition, researchers also used demographical features to study the students' performance. Those demographical features consist of gender/sex, family size, age, marital status, religion, place of birth, father occupation, mother occupation, father qualification, mother qualification, parental status, parental income status, attendance, the profile of previous education, address (urban/rural areas), college or school type, type of transport use, nationality, scholarship, internet and etc. There were a lot of demographical features that were studied by the researcher to identify the impact of those features towards the students' performance [1], [2], [7]–[9], [16], [17], [19]–[22], [28], [30], [33]–[38].

The most frequent demographical feature in previous research is gender [1]–[4], [8], [11], [12], [16]–[23], [25], [26], [28], [30], [33]–[37], [39]–[50]. However, no further investigation on the influence of gender on students' performance is discussed by those research. Even so, finding in reference [20] can be taken as information for future research. In the experiment, the researcher conducted attribute selection to remove the irrelevant features, which caused only 10 attributes to be selected, and none of them is gender attribute. This indicates that gender has no significant affect towards students' performance. This findings is supported by [40]. Despite that, researchers such as [16] and [2] found different discoveries on the gender attribute. These research stated that a few attributes had been removed during the data pre-processing phase and there was gender attribute in the final dataset. The attributes that were removed during the attribute selection were birth of place and place of living. There was also a study that stated that female students had a better academic performance in all courses compared to males [37]. Thus, these findings can be deducted that some researchers considered gender attribute as an important attribute and some of them consider that gender has lack contribution toward academic performance. There are 35 from 89 research papers that used gender as feature. The finding of gender feature on the students' performance is potentially for further investigation due to the fact that references' result are varied. Also, the findings is lacking of significant conclusion on the gender discussion.

Besides, age is also one of the features that study the students' performance [1], [2], [4], [8], [11], [16], [18], [19], [22], [33], [37], [41], [44], [45], [50]–[52]. The influence of this feature towards students' performance are not widely discussed in the present research. Nonetheless, there was a research in 2020 [53] stating that the probability of student dropout from universities was higher in the older ages compared to the young ones. This is supported by another research [54]. Although it is not directly mentioning the effect on the students' performance, the decrease in the academic might be the reason for the student dropout. Another research [41] defined that age was related to cognitive development and maturity and with increasing age, there will be developmental changes. However, findings in the research also stated that although the result showed there was a positive correlation between the feature and the academic performance, the feature did not significantly influence the students' performance [41]. As a conclusion, there are 17 research papers that used age feature to study the students' performance. The growth in the cognitive development and maturity of students might be the reasons for researchers to use the feature for students' performance research.

Some researchers used parents' occupation as one of the features [1], [4], [7], [8], [18]–[20], [22], [28], [36], [37], [42], [55], [56]. Finding in [28] showed that father's occupation was categorized as one of the influence attributes in the research despite the features undergo feature selection process using 5 different types of feature selection algorithms. This finding also was supported by [20] stating that the feature plays a major role in predicting students' grade. The parents' occupation affects the students' performance by providing financial support, motivation and basic necessities for the students' education [55], [56]. In another research, the finding showed the opposite result where fathers' occupation had no correlation with academic achievements [42]. Despite that, 13 papers claimed that parents' occupation directly impacted the student's performance due to financial support, motivation and basic necessities that are provided to them.

Other than that, there is also an influence of marital status towards students' performance, as stated in [18], [22], [28], [37], [43], [44]. Remarkably, the CGPA of married students were considerably higher than single students. The statement is supported by research [43] stating that married students tended to have mature personalities and higher responsibilities than single students. In addition, married students' social relationships and time management were also proven to be more manageable than unmarried as they had their own goals to pursue. Another researcher stated that the focus and seriousness of married students towards their studies are more significant because the marriage responsibilities and the skill of managing their time are outstanding as they needed to play a few characters of being a spouse, a parent and a student [43]. However, this is not supported by research [44] as the findings found that there was no significant difference between marital status and students' performance. The positive changes in married students such as their responsibilities, time management, maturity, focus, and seriousness influence their performance in their studies.

B. Psychometric Features

Students' performance is also affected by psychometric factors. These factors are generally towards the students' behaviour and mental development during their studies [15], [18], [21], [41], [57].

A student's interest in studying the courses was one of the factors mentioned by reference [1], [3], [4], [17], [18], [35], [37]–[39], [49], [50], [58]. The interest caused the students to develop their study strategies, high efforts and self-esteem [58]. Another research [4], [38] stated that interest, engagement time, belief and family support could increase the students' performances. There are 12 research papers that use students' interest as one of the features to study their performance. This shows that the students' interest in their studies can affect their performance.

Additionally, attendance of students during classes also can define the students' determination towards their studies. A few research use the attendance attribute to become one of the factor to affect the students' performance [1], [3], [4], [9], [10], [12], [17], [19], [21], [23], [24], [28], [33]–[35], [39], [45]–[48], [50], [59]–[63]. Research [45] stated that the correlation between students' attendance in classes and exam success was significant and positive. Attendance is also said to be a motivator for students as it is related to learning outcomes. According to [47], attendance attributes have a big impact on all components of students' performance. A higher rate of attendance is reflected in higher grades in-class activities and tests. Another impact of having a student who was continuously present in the classroom or course will develop a better understanding of the courses [64]. Thus, this feature can be used to study the students' performance as the attendance will improve their understanding of the subject.

C. Academic Features

Graduate Point Average (GPA) is the most frequent academic feature used by many researchers when studying the students' performance [1], [7], [18], [30], [33], [37], [40], [44], [48], [50], [52], [61], [65]–[69] GPA is the main indicator of students' performance, and it is usually measured on a scale with a specific range depends on the academic institution [37]. The researchers used this feature due to the tangible value that the feature has for future educational and career mobility [15]. It can also be seen as a measure of achievement in academic potential [15]. A study in [48] also used it as a benchmark to indicate whether the student can graduate on time without having an extended semester. Students' GPA is generally utilised as a predictor or dependent variable, with GPA assigned as an output or result [48]. The value of the GPA feature usually applies using numerical numbers. Some studies [67]–[69] used a range of numbers to represent the GPA ranking as a categorical variable. For example, 3.50 to 4.00 is the highest category, 3.00 to 3.50 is the middle category, etc. Nevertheless, this feature can only conclude the students' performance in their studies but not the characteristics or factors that influence the changes in the performance.

Apart from this, academic attributes such as lab work, assignments, quizzes, materials, etc. were also used by the researcher to study the students' performance [1], [3], [4], [9], [10], [17], [18], [21], [24], [26], [29], [30], [34], [48]–[50], [57], [61], [62], [64], [65], [70]–[73]. These attributes are considered internal assessment [49]. External assessment is marks achieved by the students in the final examination. A study stated that internal assessment and study behaviour could influence the students' performance [57]. This study is further proven by [63] when the students tended to be successful when they thoroughly studied the material and did homework given by the teacher. This implies that by giving assessments to the students, there will be positive effects on the students' performance and thus, this feature is suitable for students' performance research.

Despite all the features above, different data mining methods have different results in determining which features affect the students' performance. It was proven from a study [15] when a few classifiers were used to predict students' performance. Each method showed different attributes that influence the student performance, although the dataset used was identical. Hence, every data mining algorithm has its speciality, advantages and drawbacks when dealing with the data. Thus, these methods will be discussed thoroughly in the next section of this review.

IV. CLASSIFIERS

The literature on data mining has highlighted several techniques to obtain information regarding what they want to find out or investigate. Data Mining is defined as analysing data processes from different angles and summarising outcomes into useful information [31]. The term data mining is also described as determining the development of patterns in massive numbers of information, providing a huge amount of techniques and tools for analysing the data in various fields [2]. Educational Data Mining (EDM) is a term for data mining used in the educational sector; and are mostly related to education such as students' performance, behaviour and more. Data classification is an important data mining technique used in creating models that describe important data classes [31]. There are few classification algorithms depicted in Table 2, which the researchers in Educational Data Mining usually use. The classification algorithms are Decision Tree, Bayesian Classifier, K-Nearest Neighbor (KNN), Rule Learners, Random Forest, K-means, Artificial Neural Network (ANN) and Support Vector Machine (SVM).

A. Decision Tree

Decision Tree is like a flow-chart tree structure with its internal node is represented by rectangles and ovals represents the leaf nodes [32]. All internal nodes comprise two or more children nodes, and the internal nodes split to test the value of an expression of the attributes [22]. According to D. Kabakchieva, Decision Tree is defined as a classifier that generates models in tree-like structure form, starting from root attributes and ending with leaf nodes, describing the relationship among attributes and the relative importance of attributes [16] [2].

TABLE II
EDM CLASSIFICATION ALGORITHM

Classification Algorithm	Techniques	References
Decision Tree	flow-chart tree structure	[2], [3], [12], [14]–[16], [19]–[24],[4], [28], [30]–[32], [35], [48], [52], [60], [68], [69], [5], [70], [73]–[76], [6]–[11]
Bayesian Classifier	statistical type classifier	[3], [4], [20], [21], [23], [28], [30], [35], [36], [48], [69], [70], [8], [73]–[75], [77], [78], [9]–[11],[14]–[16], [19]
K-Nearest Neighbor (KNN)	a non-parametric classifying method instances	[2], [4], [60], [73], [74], [7], [9], [11], [12], [15], [16], [29], [35]
Rule Learners	two rule learner classifier : OneR JRip algorithm	[2], [4], [8], [9], [16], [69]
Random Forest	multiple Decision Trees	[3], [12], [21], [28], [33], [35], [69], [70], [73], [77]
Artificial Neural Network (ANN)	Multi-Layer Perception	[2], [4], [48], [52], [66], [70], [74], [75], [79], [10], [11], [14], [15], [20], [30], [35], [39]
Support Vector Machine (SVM)	supervised learning algorithm	[7], [11], [39], [48], [60], [66], [69], [70], [74], [76], [78], [79], [12], [80], [81], [14], [15], [19], [20], [33], [35], [38]

Each branch represents a decision (rule) while each leaf node (terminal node) corresponds to a class label (categorical or continuous value) [23]. There are three basic splitting criteria to select an attribute as a splitting point: information gain, gain ratio and gain index [27]. Two algorithms are commonly used in research which are the ID3 and J48 algorithms (the successor of the ID3 algorithm). This classifier usually is used for classification and prediction [24]. There are a few advantages of using Decision Tree classifier. The simplicity and comprehensibility of this classifier in uncovering and predicting the large or small data structure make the researchers readily understand and interpret them easily different from using Neural Network [24]. Despite the advantages that Decision Tree has, there are also some drawbacks in using this classifier. Research conducted in reference [82] stated that the learning of decision tree algorithms might cause the inability for the globally optimal decision tree to return [82]. According to [27], the time for building a tree may be higher than other classifiers, although the Decision Tree classifies quickly. Another major disadvantage is that as the number of classes increases, the classifier will suffer from a severe problem of errors propagating throughout a tree.

Decision Tree classifier is often used by many researchers because of its simplicity. In this review, the classifier has the highest number of research papers on students' performance compared to other classifiers: 35 research papers [2]–[12], [14]–[16], [19]–[24], [28], [30]–[32], [35], [48], [52], [60], [68]–[70], [73]–[76]. Although Decision Tree is widely used by researchers, the performance of the classifier is not assured. According to [74], the use of this classifier had a 98.86% maximum accuracy and 56.25% minimum accuracy. The result of the classifier's accuracy is quite the same as Bayesian classifier but with slightly higher minimum accuracy. This shows that Decision Tree can produce variety of accuracy outcomes. There were a few research that had below 70% accuracy [20], [23], [32], [35], [69], [73]. However, research such as [10], [15], [22], [52] showed good results with above 90% accuracy. Another finding of using Decision Tree classifier was when the dataset size is below 200, the classifier tended to perform better [19]. This was supported by [21] that showed increasing accuracy when the dataset increased to 200. It also showed decreasing accuracy when the dataset size was larger than 200 [19]. Decision Tree classifier also performed well when feature selection was applied to the dataset [8] [11]. The simplicity of the Decision Tree classifier will always be an attraction to the researchers to be used in their research but there is no certain that the simplicity of the classifier will produce good outcomes.

B. Bayesian Classifier

A considerable amount of literature has been published on the Bayesian Classifier. Bayesian classifier is a statistical type classifier that uses probabilities to predict class membership. There are two types of algorithms, which consist of Naïve Bayes and Bayes Networks. Naïve Bayes algorithm implies the result that the feature plays on a specific class is unaffected by the values of other features [16], [21]. It also categorises the instances based on the independent impact of each feature on classification. Bayes Networks are graphical models that are capable of describing joint conditional probability distributions [16]. During application, dependencies often exist among attributes that differ from the assumption made by Naïve Bayes so the existence of the Bayes Network is a counter problem for that situation [16]. This algorithm has high accuracy when handling large databases and delivers computational time-less than better speed [28]. Bayesian Classifier possesses simplicity, effective computation, and outstanding performance [16], [77]. This classifier can also train and evaluate faster and gives high accuracy in many domains [16]. In addition, it requires low processing memory and is also computationally inexpensive [77], [83]. The disadvantages of having this type of classifier were the strong assumptions on the feature independence and low performance in large datasets [84].

For Bayesian Classifier, many literatures highlighted the advantages of having low memory and processing power requirements, simplicity, and the ability to train and evaluate faster, making the researchers easy to implement. There are a few researchers that used Bayesian classifier in studying the students' performance classification [3], [4], [8]–[11], [14]–[16], [19]–[21], [23], [28], [30], [35], [36], [48], [69], [70], [73]–[75], [77], [78]. According to [74], Bayesian classifier had a maximum and minimum accuracy of 91.57% and 50% respectively. The large gap between the max and min accuracy indicates that the use of the classifier has the possibility to produce different results when applied to any research. In this review, 25 research papers were found using Bayesian classifier. There were only 3 research that gave promising results with above 90% classification accuracy when using the classifier: research [10], [69], [73] with 91%, 92%, 92.3% accuracy respectively. Results in previous research such as [3], [8], [30], [9], [14]–[16], [20], [21], [23], [28] showed accuracy below 76%. Another finding stated the use of Bayesian classifier was slightly improved when feature selection was applied in the research [8], [11]. Other than that, the classifier showed increasing accuracy when the dataset size was about 200 to 300. If the dataset size is below 200, the performance of Bayesian declined as shown in research [19] and [21]. Bayesian classifier is widely used among the researchers. However, the application of classifier on students' performance research is not desirable due to low classification accuracy. Future research are recommended to apply feature selection during pre-processing stage and the research' dataset size needs to be larger than 200 when using Bayesian classifier because these recommendation will improve the performance of the classifier.

C. K-Nearest Neighbour

K-Nearest Neighbour (KNN) is one of the classifiers used by the researcher. KNN is a non-parametric classification method that measures the distance between the classified instance and the closest training examples in the feature space [2], [16]. It is also defined KNN as a non-parametric method of classification. It is also called as instance-based or lazy learning algorithm because the data sample can be assigned a class label by most of the nearest neighbours [29]. This classifier aims to assign to an unseen point of the dominant class among its KNN within the training set [85]. The use of this classifier causes the research easy to understand and to implement classification technique [27], [83], [86]. KNN was also suitable for multi-modal classes and applications with many class labels [27]. The downsides of having KNN as a classifier were its complexity and the users' difficulties in understanding and interpreting the model [2]. KNN can be severely degraded if noisy or irrelevant features and if the feature scales are inconsistent with the performance [16]. KNN performance is easily affected by single training sample due to its complexity of computing when the sample similarity is huge.

This classifier does not generate the model of classification since it is a lazy learning method [82]. In terms of performance, KNN method depends on the number of dimensions. Having large data and high dimensional data will result in slower performance [84], [86]. According to [27] and [84], this classifier was considered to have high computation cost and computationally rigorous when handling increasing size of training sets.

KNN is also known as non-parametric lazy algorithms as it does not require training to execute generalisation. The use of this method is easy to understand and implement for any research. However, the drawbacks of this method are its sensitivity towards noise and irrelevant features in research. According to [85] and [86], KNN was also sensitive to the curse of dimensionality and had slow performance when handling large volumes of data. By referring to [85], this method was efficient when using low dimensional feature vectors. Thus, any future research using this method needs to perform feature selection or reduction on its dataset to avoid the problem. There were several amounts of research using KNN as a classifier in predicting student performance.

Some researchers used KNN as a classifier for their students' performance research [2], [4], [7], [9], [11], [12], [15], [16], [29], [35], [74], [60], [73]. A survey conducted on students' performance prediction from 2012 to 2016 found that only 2 research papers used KNN from 16 papers [4]. One of the papers resulted in 100% classification accuracy. Another research [7] showed that KNN could achieve 92% accuracy when predicting students' performance. Research such as [11] and [15] also achieved accuracy with above 80%. This findings were supported by research [74] where the researcher stated that KNN had a maximum accuracy of 83%. Despite that, the researcher also discovered that KNN also had 69% minimum accuracy when studying students' performance. There were few findings in some research that approved the statement such as [2], [9], [12], [16], [35], [73]. Research [2] and [12] achieved slightly higher accuracy than 69%: 70.49% and 71%. Other than that, the result of using KNN classifier in students' performance research are below 69% minimum accuracy [9], [16], [35], [73]. The use of the KNN classifier in students' performance application is not suitable as the results from the previous research are not reassuring. The accuracy of KNN classifier depends on the value of k. This is supported by research [16] stating that k-value was slightly better when the value decreased. Thus, the application of KNN classifier can be improved if the k-value is thoroughly explored.

D. Two Rule Leaner Classifier

There are also two rule learner classifiers used by the researcher in Data Mining: One Rule classifier (OneR) and JRip algorithm. OneR uses a one level-Decision Tree expressed in the form of rules set that all test one particular attribute which is the minimum-classification error attribute for prediction [2]. It is simple, cheap and always produces good rules with great accuracy for describing the structure in data [16], [2]. Also, this algorithm's is employed for the baseline comparison between other classification models and as an indicator of the particular attributes' predictive power. The JRip algorithm implements the Repeated Instrumental Pruning to Produce Error Reduction (RIPPER) algorithm. This algorithm was used to examine the classes in increasing size and an initial set of rules is used for class generation by using the incremental reduced-error pruning [16].

The researcher rarely uses OneR and JRip algorithms. Most of the researchers utilised D. Kabakchieva technique to gain information on these algorithms. By referring to [16], it was found that JRip rule learner performs slightly better with 63% overall accuracy than the OneR classifier, 54%-55% while [17] stated that the OneR classifier has the least accurate compared to the other three classifiers. These findings demonstrated the student performance prediction which the technique was less likely recommended due to the fact that the findings produced inaccurate classification performance as referred to the previous research.

The used of Rule Learner classifier in student performance application are not widely discussed. Some research such as [2],[4],[8],[9],[16],[69] used this classifier for the application, and some of them also produced good results. According to [4], a literature survey was done on students' performance prediction from 2012 to 2016. From 16 research papers, there were 9 papers that used Rule Learner as a classifier. The survey's finding also stated that this classifier's use had the lowest average classification accuracy of 75.85%. The finding was also supported by research such as [8] and [9], OneR achieved the highest classification accuracy with 76.73% and JRip scored the third highest overall classification accuracy with slightly above 70% respectively. The highest classification accuracy recorded from using rule learner classifier was 96.7% [4]. The use of Rule Learner classifier in students' performance research is not suitable as the previous studies showed low classification accuracy results below 65% [2] [16]. However, it can be further re-examined and discussed by the researchers as there was a research that result in high classification accuracy when using the Rule Learner classifier [4].

E. Random Forest

Random Forest is another well-known classifier the researchers adopted in their work. This classifier builds multiple Decision Trees for the given data and it predicts the class label by taking majority votes of the Decision Trees for the test sample [21]. According to [73], Random Forest is defined as a supervised ensemble machine learning approach that operates by constructing some Decision Trees and producing as its output is made of individual trees classes. When comparing to Decision Trees where each node is split using the best among attributes, each node of Random Forest is split using the best among a subset of predictors randomly chosen at the node [73]**.** Random Forest has its advantages when using it as a method for data classification. This classifier was said to be easy to interpret and understand and it was also non-parametric. Thus, the linearity of the input data set will not be a hindrance [77]. Another advantage of this classifier was the pruning of the trees will be unnecessary if the parameters are available and easily entered [77]. In addition, the classification model was fast and scalable and robust to irrelevant text present in a document [77]. The obvious drawback of having Random Forest as a classifier is that it is easily overfitting its class. However, this can be prevented by reducing the trees number in the classifier and decreasing the present vague links [77].

There were a few previous research performed using Random Forest classifier when studying students' performance [3], [12], [21], [28], [33], [35], [69], [70], [73], [77]. Some of this research also produced good results in the classification accuracy when using the classifier [3],[12],[21],[28],[33]. According to [3], a research was conducted to make comparative study on marks prediction using different data mining techniques. A few classifiers were used in this study consisted of Decision Tree, Random Forest, Naïve Bayes, Naïve Bayes Multinomial, K-star and IBK. The resulted showed that Random Forest classifier managed to get the highest classification accuracy with 76.67%. Another research [12] also showed that Random Forest gave an optimum accuracy (90%) when predicting students' performance using Learning Management System (LMS) data. The research such as [28] and [33] were also showed that Random Forest was capable of achieving above optimum accuracy with 99% and 94.45% respectively. A finding from [21] stated that Random Forest could get good results with increasing dataset size. However, the used of this classifier was time-consuming compared to other classifiers in the study [21]. Previous research that used Random Forest classifier produced promising results when studying students' performance. The only drawbacks of using this classifier is the time-consuming factor as stated in [21]. However, the time in the research was recorded in millisecond. Thus, the time-consuming factor is not a major problem in conducting research when the classifier is capable of delivering contribution due to the high accuracy results as shown in the previous studies.

F. Artificial Neural Network

Artificial Neural Network (ANN) is used to produce classification models in the mathematical model form, consisting of interconnected computational elements (neurons) and processing information using a connectionist approach to computation [2]. ANN has an algorithm that is called as Multi-Layer Perception. According to [87], Multi-Layer Perception is defined as a classifier in which the network's weights can be determined by solving a problem of quadratic programming with linear constraints. It is different from the traditional neural network training that solves a non-convex, unconstrained minimisation problem. It is generally used for learning from training batch instances by repetitively running the algorithm across the training set until a prediction vector is discovered that is correct over the whole training set. This prediction rule is then used for predicting the labels on the test set [87]. The advantage of having ANN as a classifier is that it can detect all possible interactions between predictors' variables and run complete detection without having any uncertainty even in a complex non-linear relationship between dependent and independent variables [15]. The disadvantage that this classifier possessed is the sensitivity towards overtraining, especially such noisy and non-stationary data. These networks have difficulty interpreting the correlation between independent and dependent variables [85] [34].

ANN is said to have the ability to identify interactions between predictors' variables which makes it suitable for analysing any prediction especially student performance. The complexity and difficulty in understanding and interpreting the method making it a non-user-friendly classifier.

ANN is one of the classifiers that often used by many researchers. There are also many application of NN classifier in students' performance research such as [2], [4], [10], [11], [14], [15], [20], [30], [35], [39], [48], [52], [66], [70], [74], [75], [79]. A survey [4] stated that NN had an average of 78.7% accuracy. A research [14] also achieved an accuracy of 79.22% when predicting students' performance. Despite that, the used of ANN classifier in some research showed that the classifier was capable of producing above 90% classification accuracy [10], [15], [39], [52], [74], [79]. A graphical representation from a review showed that ANN could achieve a maximum accuracy of 98% and a minimum accuracy of 62.5% [74]. The finding was also supported by another review [15] that discovered 2 research paper from 2012 and 2013 that achieved accuracy of 98% and 97% respectively. Other than that, ANN can also perform well when applying feature selection to the dataset. A research showed that the classifier's performance increased from 74% to 81% accuracy when feature selection was applied and the changes in the accuracy were large compared to other classifiers used in the study [11]. Thus, the promising results from previous research make ANN classifier suitable to be used in students' performance classification.

G. Support Vector Machine

Another classifier for Data Mining is Support Vector Machine (SVM). SVM is a supervised learning algorithm that is used for classification, regression and outlier detection [35], [50], [60]. Vapnik in 1995 and his group AT&T Bell Laboratories proposed this classification and regression technique in [66], [88]–[91]. SVM classifier's goal is to find a linear hyperplane or "thickest hyperplane" that is also known as decision boundary that separates the data in such a way that the margin is maximised [60], [66], [88], [75]. Choosing the boundary that maximises the margin will decrease the chances for misclassification of unknown items in the future [88], [90]. SVM is considered a good classifier because of its high generalisation performance without adding a priori knowledge even when the input space dimension is very high [89]. If the data are not linearly separable, the researcher can choose a few kernels function to adapt to the situation, consisting of kernel functions from an array of linear, radial basis, sigmoid second-order multiple, polynomial and reverse second-order kernel [17]. The idea of SVM is to find the points of data known as support vectors, which define the widest linear margin between two classes. Two tricks can be performed to the non-linear class boundaries; first, mapping the data to a higher dimension, where there is linear boundary and second, allowing misclassification by defining a soft margin. A compromise of these two approaches will avoid overfitting and preserve good classification accuracy [18], [81], [92]. According to [81], if there are cases where the points are not linearly separable, SVM has a parameter, *C,* that computes the hyperplane that maximises the distances to support vectors for a given parameter setting. Other than that, the use of advanced kernel methods can transform a non-linear input space to a linear feature space, and it is suitable to be used when encounters problems that are not separated linearly.

SVM is a well-known classifier to the researchers due to its adjustable hyperparameter that can adapt to every research's situation. In this review, there are 22 research papers that used SVM as classification technique in students' performance research [7], [11], [12], [14], [15], [19], [20], [33], [35], [38], [39], [48], [60], [66], [69], [70], [74], [76], [78]–[81]. A finding was declared by a systematic review [74] stating that SVM had minimum accuracy of 80%. From 22 research papers that have been reviewed, there are few papers with slightly above 80% accuracy [11], [12], [15], [33], [69]. And only 2 research showed accuracy below 80%: research [14] with 75.28% and research [35] with 75%. Research [74] also stated that using SVM classifier could achieve a maximum accuracy of 98%.Thus, research such as [7], [39], [60], [70], [76], [78], [79] confirmed the declaration that the classifier can achieve above 90% accuracy. Another finding stated that SVM accuracy increased with increasing dataset size [19]. The result was considered a good finding because the other classifier such as Decision Tree showed increasing accuracy in the first 100 dataset size and it dropped after the dataset size increased for another 100 while another classifier in the study, Naïve Bayes showed vice versa graph [19]. Hence, the use of SVM classifier in students' performance classification is recommended as it can achieve high accuracy and has acceptable minimum accuracy. It is also reliable when dealing with increasing dataset size during the research.

It can be seen that all techniques have been applied in various area including student performance research. Every technique has its advantages and disadvantages when using them in any research. Most of the techniques have a specific tuning that needs to be adjusted to maximise their performance during classification.

V. CONCLUSION

In this article, a critical review of students' performance classification was presented focusing on its data, features that influence the students' performance and classifiers. An overview of the data properties based on its quantities, type of data used, as well as the features' type and values and also the classifier properties, including the working principle, parameters, applications as well as comparison on the performance of the classifier with other classifiers were included. Different existing methods for the data classification of the students' performance is tabulated. The studies presented were on the EDM which studied the classification of students' performance, and it is aimed to observe the correlation between the data, the features and the classifiers that affect the students' performance.

ACKNOWLEDGEMENT

The authors would also like to express gratitude for the supports given by the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam. Also, this research work is funded by Special Research Grant (GPK). Grant no 600-RMC/GPK 5/3 (039/2020).

REFERENCES

[1] A. Abu Saa, M. Al-Emran, and K. Shaalan, *Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques*, vol. 24, no. 4. Springer Netherlands, 2019.

[2] D. Kabakchieva, "Student performance prediction by using data

mining classification algorithms," *Int. J. Comput. Sci. Manag. Res.*, vol. 1, no. 4, pp. 686–690, 2012.
[3] B. Kapur, N. Ahluwalia, and S. R, "Comparative Study on Marks Prediction using Data Mining and Classification Algorithms," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 3, pp. 632–636, 2017, [Online]. Available: http://www.ijarcs.info/index.php/Ijarcs/article/view/3066.
[4] M. Kumar, A. J. Singh, and D. Handa, "Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques," *Int. J. Educ. Manag. Eng.*, vol. 7, no. 6, pp. 40–49, 2017, doi: 10.5815/ijeme.2017.06.05.
[5] D. Ph, M. Daud, and D. Ph, "Final Grade Prediction of Secondary School Student using Decision Tree," *Int. J. Comput. Appl.*, vol. 115, no. 21, pp. 32–36, 2015.
[6] S. A. Kumar, "Efficiency of Decision Trees in Predicting Student ' S Academic Performance," *Comput. Sci. Inf. Technol. (CS IT)*, pp. 335–343, 2011, doi: 10.5121/csit.2011.1230.
[7] S. Wiyono and T. Abidin, "Comparative Study of Machine Learning Knn, Svm, and Decision Tree Algorithm To Predict Student'S Performance," *Int. J. Res. -GRANTHAALAYAH*, vol. 7, no. 1, pp. 190–196, 2019, doi: 10.29121/granthaalayah.v7.i1.2019.1048.
[8] Y. K. Salal, S. M. Abdullaev, and M. Kumar, "Educational Data Mining : Student Performance Prediction in Academic," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 4c, 2019.
[9] C. Anuradha and T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance," *Indian J. Sci. Technol.*, vol. 8, no. 15, 2015.
[10] H. Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers," *Int. J. Mod. Educ. Comput. Sci.*, vol. 9, no. 8, pp. 9–15, 2017, doi: 10.5815/ijmecs.2017.08.02.
[11] A. A. Rimi, "Developing Classifier for the Prediction of Students' Performance Using Data Mining Classification Techniques," *AURUM J. Eng. Syst. Archit.*, vol. 4, no. 1, pp. 73–91, 2020.
[12] A. Dutt and M. A. Ismail, "Can We Predict Student Learning Performance from LMS Data? A Classification Approach," *Adv. Soc. Sci. Educ. Humanit. Res.*, vol. 326, no. Iccie 2018, pp. 24–29, 2019, doi: 10.2991/iccie-18.2019.5.
[13] S. A.SAlloum, M. Alshurideh, Elnagar, Ashraf, and K. Shaalan, *Mining in Educational Data: Review and Future Directions*, vol. 2, no. 1153. Springer International Publishing, 2020.
[14] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
[15] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015, doi: 10.1016/j.procs.2015.12.157.
[16] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," *Cybern. Inf. Technol.*, vol. 13, no. 1, pp. 61–72, 2013, doi: 10.2478/cait-2013-0006.
[17] E. Filiz and E. Öz, "Finding the best algorithms and effective factors in classification of Turkish science student success," *J. Balt. Sci. Educ.*, vol. 18, no. 2, pp. 239–253, 2019, doi: 10.33225/jbse/19.18.239.
[18] A. O. Ameen, M. A. Alarape, and K. S. Adewole, "Students' Academic Performance and Dropout Prediction," *Malaysian J. Comput.*, vol. 4, no. 2, p. 278, 2019, doi: 10.24191/mjoc.v4i2.6701.
[19] S. Urkude and K. Gupta, "Student intervention system using machine learning techniques," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special Issue 3, pp. 2061–2065, 2019, doi: 10.35940/ijeat.F1392.0986S319.
[20] V. Ramesh, "Predicting Student Performance : A Statistical and Data Mining Approach," *Int. J. Comput. Appl. (0975 – 8887)*, vol. 63, no. 8, pp. 35–39, 2013.
[21] K. Prasada, M. V. P. Chandra, and B. Ramesh, "Predicting Learning Behavior of Students using Classification Techniques," *Int. J. Comput. Appl.*, vol. 139, no. 7, pp. 15–19, 2016, doi: 10.5120/ijca2016909188.
[22] Y. K. Saheed, T. O. Oladele, A. O. Akanni, and W. M. Ibrahim, "Student performance prediction based on data mining classification techniques," *Niger. J. Technol.*, vol. 37, no. 4, p. 1087, 2018, doi: 10.4314/njt.v37i4.31.
[23] K. Yadav and R. Thareja, "Comparing the Performance of Naive Bayes And Decision Tree Classification Using R," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 12, pp. 11–19, 2019, doi: 10.5815/ijisa.2019.12.02.
[24] K. Sunday, P. Ocheja, S. Hussain, S. S. Oyelere, O. S. Balogun, and F. J. Agbo, "Analyzing student performance in programming education using classification techniques," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 2, pp. 127–144, 2020, doi: 10.3991/ijet.v15i02.11527.
[25] S. Ali, Z. Haider, F. Munir, H. Khan, and A. Ahmed, "Factors Contributing to the Students Academic Performance: A Case Study of Islamia University Sub-Campus," *Am. J. Educ. Res.*, vol. 1, no. 8, pp. 283–289, 2013, doi: 10.12691/education-1-8-3.
[26] A. Althnian *et al.*, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Appl. Sci.*, vol. 11, no. 2, pp. 1–18, 2021, doi: 10.3390/app11020796.
[27] S. S. Ahmed and D. L. Abd AL-Nabi, "Survey on Classification Algorithms for Data Mining :(Comparison and Evaluation)," vol. 4, no. 8, pp. 18–25, 2013.
[28] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018, doi: 10.11591/ijeecs.v9.i2.pp447-459.
[29] A. S. Nagesh, C. V. S. Satyamurty, and K. Akhila, "Predicting Student Performance using KNN Classification in Bigdata Environment," *CVR J. Sci. Technol.*, vol. 13, no. December, pp. 83–87, 2017.
[30] M. Osmanbegovic, Edin; Suljic and Article, "Data Mining Approach for Predicting Student Performance," *Econ. Rev. J. Econ. Bus.*, vol. 10, no. 1, p. 12, 2012.
[31] K. Sumathi, S. Kannan, and K. Nagarajan, "Data Mining: Analysis of student database using Classification Techniques," *Int. J. Comput. Appl.*, vol. 141, no. 8, pp. 22–27, 2016, doi: 10.5120/ijca2016909703.
[32] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," pp. 334–337, 2013.
[33] M. Laxmi, I. Raju, G. Latha Pavani, and K. Vijaya Kumar, "ANALYSIS OF STUDENT ACADEMIC PERFORMANCE LEADING TECHNOLOGY USING CLASSIFICATION ALGORITHMS," *J. Crit. Rev.*, vol. 7, no. 19, pp. 7251–7259, 2020.
[34] Ş. Aydoğdu, "Predicting student final performance using artificial neural networks in online learning environments," *Educ. Inf. Technol.*, vol. 25, no. 3, pp. 1913–1927, 2020, doi: 10.1007/s10639-019-10053-x.
[35] V. Vijayalakshmi and K. Venkatachalapathy, "Comparison of Predicting Student's Performance using Machine Learning Algorithms," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 12, pp. 34–45, 2019, doi: 10.5815/ijisa.2019.12.04.
[36] B. K. Bhardwaj, "Data Mining : A prediction for performance improvement using classification," *Int. J. Comput. Sci. Inf. Secur.*, vol. 9, no. 4, 2011.
[37] R. O. Salem, N. Al-Mously, N. M. Nabil, A. H. Al-Zalabani, A. F. Al-Dhawi, and N. Al-Hamdan, "Academic and socio-demographic factors influencing students' performance in a new Saudi medical school," *Med. Teach.*, vol. 35, no. SUPPL. 1, pp. 83–89, 2013, doi: 10.3109/0142159X.2013.765551.
[38] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, and E. Wani, "Prediction of Student Academic Performance By an Application of Data Mining Techniques," *Manag. Artif. Intell.*, vol. 6, no. January, pp. 110–114, 2011.
[39] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation," *Soft Comput.*, vol. 24, no. 24, pp. 18477–18487, 2020, doi: 10.1007/s00500-020-05075-4.
[40] A. M. Remali, M. A. Ghazali, M. K. Kamaruddin, and Y. K. Tan, "Understanding academic performance based on demographic factors, motivation factors and learning styles," *Int. J. Asian Soc. Sci.*, vol. 3, no. 9, pp. 1938–1951, 2013, [Online]. Available: http://www.aessweb.com/pdf-files/Ijass-si-3(9)-1938-1951.pdf.
[41] E. E. Ebenuwa-Okoh, "Influence of Age, Financial Status, and Gender on Academic Performance among Undergraduates," *J. Psychol.*, vol. 1, no. 2, pp. 99–103, 2010, doi: 10.1080/09764224.2010.11885451.
[42] G. A. Postiglione, X. Ailei, J. Jung, and H. Yanbi, "Rural Students in a Chinese Top-Tier University: Family Background, School Effects, and Academic Performance," *Chinese Educ. Soc.*, vol. 50, no. 2, pp. 63–74, 2017, doi: 10.1080/10611932.2017.1326774.
[43] J. Thomas, M. Raynor, and A. Al-Marzooqi, "Marital status and

gender as predictors of undergraduate academic performance: a United Arab Emirates context," *Learn. Teach. High. Educ. Gulf Perspect.*, vol. 9, no. 2, pp. 37–45, 2012, doi: 10.18538/lthe.v9.n2.80.

[44] R. Tabassum and N. Akhter, "Effect of Demographic Factors on Academic Performance of University Students," *J. Res. Reflections Educ.*, vol. 14, no. 1, pp. 64–80, 2020.

[45] A. Lukkarinen, P. Koivukangas, and T. Seppälä, "Relationship between Class Attendance and Student Performance," *Procedia - Soc. Behav. Sci.*, vol. 228, no. June, pp. 341–347, 2016, doi: 10.1016/j.sbspro.2016.07.051.

[46] V. Kassarnig, E. Mones, A. Bjerre-Nielsen, P. Sapiezynski, D. D. Lassen, and S. Lehmann, "Academic performance and behavioral patterns," *EPJ Data Sci.*, vol. 7, no. 1, pp. 1–16, 2018, doi: 10.1140/epjds/s13688-018-0138-8.

[47] J. M. N. Jover and J. A. M. Ramírez, "Academic performance, class attendance and seating location of university students in practical lecture," *J. Technol. Sci. Educ.*, vol. 8, no. 4, pp. 337–345, 2018, doi: 10.3926/jotse.353.

[48] M. Asiah, K. Nik Zulkarnaen, D. Safaai, M. Y. Nik Nurul Hafzan, M. Mohd Saberi, and S. Siti Syuhaida, "A Review on Predictive Modeling Technique for Student Academic Performance Monitoring," *MATEC Web Conf.*, vol. 255, p. 03004, 2019, doi: 10.1051/matecconf/201925503004.

[49] D. Magdalene Delighta Angeline, "Association Rule Generation for Student Performance Analysis using Apriori Algorithm," *SIJ Trans. Comput. Sci. Eng. its Appl.*, vol. 1, no. 1, pp. 12–16, 2013.

[50] M. Kumar and Y. K. Salal, "Systematic review of predicting student's performance in academics," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 54–61, 2019, doi: 10.13140/RG.2.2.26667.69923.

[51] Z. Ren and M. Sweeney, "Predicting Student Performance Using Personalized Analytics," *Computer (Long. Beach. Calif).*, vol. 49, no. 4, pp. 61–69, 2016, doi: 10.1109/MC.2016.119.

[52] E. N. Ogor and C. Islands, "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques Department of Natural Sciences Turks & Caicos Islands Community College Visualization and Articulation," *Fourth Congr. Electron. Robot. Automot. Mech.*, pp. 0–5, 2007, doi: 10.1109/CERMA.2007.78.

[53] B. M. Kehm, M. R. Larsen, and H. B. Sommersel, "Student dropout from universities in Europe: A review of empirical literature," *Hungarian Educ. Res. J.*, vol. 9, no. 2, pp. 147–164, 2020, doi: 10.1556/063.9.2019.1.18.

[54] M. Søgaard *et al.*, *Clearinghouse – research series A systematic review Danish Clearinghouse for Educational Research*, no. 15. 2013.

[55] O. longinus Chukwudi, "Influence of Parental Occupation and Parental Level of Education on Students ' Academic Performance in," *Int. J. Res. Innov. Soc. Sci. |Volume*, vol. II, no. XII, pp. 201–211, 2018, [Online]. Available: https://www.rsisinternational.org/journals/ijriss/Digital-Library/volume-2-issue-12/201-211.pdf.

[56] F. U. Odikpo, "Influence of Parental Occupation and Educational Level on Academic Achievement of Senior Secondary School Students in Anambra State," *J. Educ. Res. Dev.*, vol. 4, no. 1, pp. 167–175, 2021.

[57] A. Khan and S. K. Ghosh, *Student performance analysis and prediction in classroom learning: A review of educational data mining studies*, vol. 26, no. 1. Education and Information Technologies, 2021.

[58] A. J. Martin, H. W. Marsh, D. M. Mclnerney, J. Green, and M. Dowson, "Getting Along with Teachers and Parents: The Yields of Good Relationships for Students' Achievement Motivation and Self-Esteem," *Aust. J. Guid. Couns.*, vol. 17, no. 2, pp. 109–125, 2007, doi: 10.1375/ajgc.17.2.109.

[59] V. Kassarnig, A. Bjerre-Nielsen, E. Mones, S. Lehmann, and D. D. Lassen, "Class attendance, peer similarity, and academic performance in a large field study," *PLoS ONE*, vol. 12, no. 11. 2017, doi: 10.1371/journal.pone.0187078.

[60] S. A. Oloruntoba, "Student Academic Performance Prediction Using Support Vector Machine," *Ijesrt Int. J. Eng. Sci. Res. Technol.*, vol. 6, no. 12, pp. 588–598, 2017.

[61] M. Tiwari, R. Singh, and N. Vimal, "An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education," *Int. J. Comput. Sci. Mob. Comput.*, vol. 2, no. 2, pp. 53 – 57, 2013.

[62] S. Rajesh Suyal and M. Mukund Mohod, "Quality IMPROVISATION OF STUDENT PERFORMANCE USING DATA MINING TECHNIQUES," *Int. J. Sci. Res. Publ.*, vol. 4, no. 1, pp. 2250–3153, 2014, [Online]. Available: www.ijsrp.org.

[63] J. -P. Vandamme, N. Meskens, and J. -F. Superby, "Predicting Academic Performance by Data Mining Methods," *Educ. Econ.*, vol. 15, no. 4, pp. 405–419, 2007, doi: 10.1080/09645290701409939.

[64] N. Hajizadeh and M. Ahmadzadeh, "Analysis of factors that affect the students academic performance - Data Mining Approach," *Int. J. Adv. Stud. Comput. Sci. Eng. IJASCSE*, vol. 3, no. 8, pp. 1–4, 2014, [Online]. Available: http://arxiv.org/abs/1409.2222.

[65] M. Shotwell, C. H. Apigian, M. Shotwell, and C. H. Apigian, "Student Performance and Success Factors in Learning Business Statistics in Online vs . On- Ground Classes Using a Web-Based Assessment Platform Assessment Platform," *J. Stat. Educ.*, vol. 23, 2015, doi: 10.1080/10691898.2015.11889727.

[66] F. D. Kentli and Y. Sahin, "An SVM approach to predict student performance in manufacturing processes course.pdf," *Energy Educ. Sci. Technol. Part B Soc. Educ. Stud.*, vol. 3, no. 4, pp. 535–544, 2011.

[67] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch, "Predicting student performance: An application of data mining methods with an educational web-based system," *Proc. - Front. Educ. Conf.*, vol. 1, pp. 1–6, 2003.

[68] M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 528–533, 2016, doi: 10.7763/ijiet.2016.v6.745.

[69] N. Alangari and R. Alturki, "Predicting students final GPA using 15 classification algorithms," *Rom. J. Inf. Sci. Technol.*, vol. 23, no. 3, pp. 238–249, 2020.

[70] 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, doi: 10.3390/e22010012.

[71] C. Tucker, B. K. Pursel, and A. Divinsky, "Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes," *Comput. Educ. J.*, vol. 5, no. 4, pp. 84–95, 2014, doi: 10.18260/1-2--22840.

[72] R. A. Alturki, "Measuring and improving student performance in an introductory programming course," *Informatics Educ.*, vol. 15, no. 2, pp. 183–204, 2016, doi: 10.15388/infedu.2016.10.

[73] Y. Abubakar, N. Bahiah, and H. Ahmad, "Prediction of Students' Performance in E-Learning Environment Using Random Forest," *Int. J. Innov. Comput.*, vol. 7, no. 2, pp. 1–5, 2017, [Online]. Available: http://se.fsksm.utm.my/ijic/index.php/ijic.

[74] M. Kumar and Y. K. Salal, "Systematic Review of Predicting Student ' s Performance in Academics," no. 3, pp. 54–61, 2019.

[75] A. Dinesh Kumar, R. Pandi Selvam, and K. Sathesh Kumar, "Review on prediction algorithms in educational data mining," *Int. J. Pure Appl. Math.*, vol. 118, no. Special Issue 8, pp. 531–537, 2018.

[76] M. S. Mythili and A. R. M. Shanavas, "A Novel Approach to Predict the Learning Skills of Autistic Children using SVM and Decision Tree," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 6, pp. 7288–7291, 2014, [Online]. Available: www.ijcsit.com.

[77] A. Gupte, S. Joshi, P. Gadgul, and A. Kadam, "Comparative Study of Classification Algorithms used in Sentiment Analysis," vol. 5, no. 5, pp. 6261–6264, 2014.

[78] H. Wilhelmiina, "Comparison of machine learning methods for intelligent tutoring systems," pp. 525–534, 2006.

[79] A. Tekin, "Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach," *Eurasian J. Educ. Res.*, vol. 14, no. 54, pp. 207–226, 2014, doi: 10.14689/ejer.2014.54.12.

[80] C. Huang, S. Chu, and C. Guan, "Implementation and performance evaluation of parameter improvement mechanisms for intelligent e-learning systems," vol. 49, pp. 597–614, 2007, doi: 10.1016/j.compedu.2005.11.008.

[81] H. Yu, J. Han, and K. C. C. Chang, "PEBL: Positive example based learning for web page classification using SVM," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 239–248, 2002.

[82] A. H. Aliwy and E. H. A. Ameer, "Comparative Study of Five Text Classification Algorithms with their Improvements," no. November 2018, 2017.

[83] K. Kowsari, K. J. Meimandi, M. Heidarysafa, and S. Mendu, "Text Classification Algorithms : A Survey," pp. 1–68, 2019, doi: 10.3390/info10040150.

[84] S. R. Pakize, "Comparative Study of Classification Algorithms Based

On MapReduce Model," vol. 1, no. 7, pp. 251–254, 2014.
[85] F. Lotte and M. Congedo, "TOPICAL REVIEW A review of classification algorithms for EEG-based brain – computer interfaces," vol. 1, 2007, doi: 10.1088/1741-2560/4/2/R01.
[86] N. C. S. Reddy, K. S. Prasad, and A. Mounika, "Classification Algorithms on Datamining : A Study," vol. 13, no. 8, pp. 2135–2142, 2017.
[87] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijctt-v48p126.
[88] J. Khairnar and M. Kinikar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification," *Int. J. Sci. Res. Publ.*, vol. 3, no. 6, pp. 1–6, 2013, [Online]. Available: www.ijsrp.org.
[89] N. Krishnaveni and V. Radha, "Feature Selection Algorithms for Data Mining Classification : A Survey," vol. 12, no. February, 2019, doi: 10.17485/ijst/2018/v12i6/139581.
[90] M. Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016, doi: 10.1109/ACCESS.2016.2568756.
[91] C. J. Huang, S. S. Chu, and C. T. Guan, "Implementation and performance evaluation of parameter improvement mechanisms for intelligent e-learning systems," *Comput. Educ.*, vol. 49, no. 3, pp. 597–614, 2007, doi: 10.1016/j.compedu.2005.11.008.
[92] W. Hämäläinen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4053 LNCS, no. January, pp. 525–534, 2006, doi: 10.1007/11774303_52.