

Research Article

Data Mining Techniques to Analyze the Impact of Social Media on Academic Performance of High School Students

Saman Amjad,¹ Muhammad Younas^{ID},¹ Muhammad Anwar^{ID},² Qaisar Shaheen^{ID},³ Muhammad Shiraz,⁴ and Abdullah Gani^{ID}⁵

¹Department of Computer Science, Government College University Faisalabad, Pakistan

²Department of Information Science, Division of Science and Technology, University of Education, Lahore, Pakistan

³Department of Computer Science, Islamia University Bahawalpur, Rahim Yar Khan Sub Campus, Pakistan

⁴Department of Computer Science, Federal Urdu University of Arts, Science & Technology, Islamabad, Pakistan

⁵Faculty of Computing and Informatics, University Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Correspondence should be addressed to Abdullah Gani; abdullahgani@ums.edu.my

Received 13 November 2021; Revised 6 February 2022; Accepted 22 February 2022; Published 20 March 2022

Academic Editor: B. B. Gupta

Copyright © 2022 Saman Amjad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main purpose of educational institutions is to provide quality education to their students. However, it is difficult to analyze large data manually. Educational data mining is more effective as compared to statistical methods used to explore data in educational settings to analyze students' performance. The objective of the study is to use different data mining techniques and find their performance and impact of different features on students' academic performance. The dataset was collected from the Kaggle repository. To analyze the dataset, different classification algorithms were applied like decision tree, random forest, SVM classifier, SGD classifier, AdaBoost classifier, and LR classifier. This research revealed that random forest achieved a higher score (98%). The score of decision tree, AdaBoost, logistic regression, SVM, and SGD is 90%, 89%, 88%, 86%, and 84%, respectively. Results show that technology greatly influences student performance. The students who use social media throughout the week showed low performance as compared to the students who use it only at weekends. Furthermore, the impact of other features on the performance of students is also measured.

1. Introduction

Student's performance modeling is one of the challenging and popular research topics in educational data mining (EDM) [1]. Multiple factors influence the academic performance in nonlinear ways. The widespread availability of educational datasets further made educational data mining more attractive to the researchers. EDM is a field in which data mining algorithms are applied on educational data to improve and predict the performance of education in institute students [2].

Information Technology (IT) is an important part of learning process [3]. It greatly influences the online student performance and GPA. The study [4] believes that the use of technology such as internet is one of the most important

factors that can influence the educational performance of students positively or negatively. Students are spending too much time on social media sites like Facebook and do have not enough time to study. This behavior leads students toward poor performance during high school studies, and they consider themselves difficult to survive in higher studies. EDM can detect this poor behavior pattern at right time to maximize the student grades and minimize the failure rate of weak students.

Social media greatly influence the school-age students. Social media influences students' academic and personal lives. Students use social media for academic purposes to improve their performance. Teachers and students both can use social media as a teaching and learning tool for ease and improve learning and teaching process [5].

The objective of the study is to predict student performance with the use of technology, weekday-social-media-use, and weekend-social-media-use and, furthermore, to find out the student who desires to get higher education in advance and to find how parent's education influences the student performance.

In this paper, data is collected from Callboard 360 LMS (learning management system). We used six machine learning techniques (DT, SVM classifier, SGD classifier, RF classifier, AdaBoost, and LR classifier) to determine the patterns inside the student performance data.

The results show that technology greatly influences the student performance. Six classifiers are used to identify the performance on the basis of different features like romantic status, use of technology, weekday-social-media-use, weekend-social-media-use, parent education, and living area. It helps the teachers to identify the fair, good, and poor students.

The proposed work analyzes performance and finds the desirable and undesirable student behaviors of students, which will help to combine students into classes based on different performance capabilities, furthermore predicting the student's social activities.

2. Relate Work

Educational institutes face different problems to identify reasons of drop-out, graduate not on time, pass to fail ratio, effect of parent's involvement on student performance, effect of attendance on student performance, predicting student's performance on the basis of previous marks, and many more. For solving such problems, several studies present machine learning and statistical solutions.

The study [4] believes that the use of technology such as internet is one of the factors that can influence educational performance of students positively or negatively. Students are spending too much time on social media websites like Facebook and do have not enough time to study which leads students towards poor performance during high school studies and ultimately consider themselves difficult to survive in higher studies. EDM can detect this poor behavior pattern at right time to maximize the student grades and minimize the failure rate of students. Another study [5] shows that social media influences the school-age students positively. Students use social media for academic purposes to improve their performance. Teachers and students both can use social media as a teaching and learning tool for ease and improve learning and teaching process.

Study [6] used decision tree, a nonlinear classifier to generate tree and rules. For analyzing results, the J48 algorithm is used as an analyzing tool. Dataset is collected through the surveys of students of master and Ph.D. studies.

Study [7] finds the student's academic results based on cluster groups and uses standard (statistical) algorithms to collect and manage their score data corresponding to the level of their performance. The K-mean clustering algorithm is used to analyze academic performance. Another study [8] used knowledge discovery and data mining tools for extracting useful information from data repository, which is used to

enhance the quality of education. DM methods are used for decision making in educational systems. A decision tree (DT) algorithm is used for data searching based on divide and conquer rule. Student's academic performance is measured by applying decision tree algorithms on students' past academic data to predict and analyze the data. These performance measures help to find the dropout students and to identify who need special coaching and allocation of instructor for suitable advice and counseling.

Study [9] declared that decision tree is the most broadly used supervised classification algorithm. Its step creation is fast and easy. The DT classifier is applied on any field. Student qualitative dataset is collected from educational data mining. Different decision tree algorithms are applied on dataset and compared the performance. These algorithms are CART, C4.5, and ID3. The contrast outcome shows that the Gini index of CART influences the information of gain ratio of ID3 and C4.5. The CART algorithm performance and correctness are greater than ID3 and C4.5, because the DT algorithm results prove that student's performance is influenced by qualitative features.

Reference [10] declared that data mining techniques increasingly merged day by day with the educational field. Data mining and education field are combined called educational data mining that help to identify the features and information of students. This study uses to predict and analyze the performance of bachelor and master students at university level students. The performance is analyzed with two algorithms: decision tree and fuzzy genetic algorithms. The dataset contains features internal-marks, sessional-marks, and admission-marks which are used to identify the results. Internal-marks contain attendance-marks, AVG-marks, sessional-marks, and assignment-marks. Weighted marks obtain from matric and interclass. In master degree, examination marks are also included. A systematic model is used to enhance the performance of students in the early stage and in time. To find the result and solution in early stage, conduct good result in final examination. Students also view their result and new updates. Many companies connect to the educational organizations to find out students according to their needs.

Reference [11] declared that large amount of data is stored in different technological spaces and makes new data quickly and easily. Data mining is also combined with these technologies. With the help of data mining techniques, important information from ordinary data can be taken out. Because of these techniques, data can be produced quickly and easily daily or each second. Using data mining methods provide meaningful knowledge. An educational database contains huge amount of data related to student data mining methods applying on this data. This study defines how to use DM algorithms such as KNN, naïve Bayes, and DT algorithm. Apply these algorithms on student raw data and find the best result.

Reference [12] narrated that college students have great facility of internet. Internet educates the students in living and learning process. This study discloses the connection between internet use behavior and educational performance of students. It also analyzes students that are undergraduates

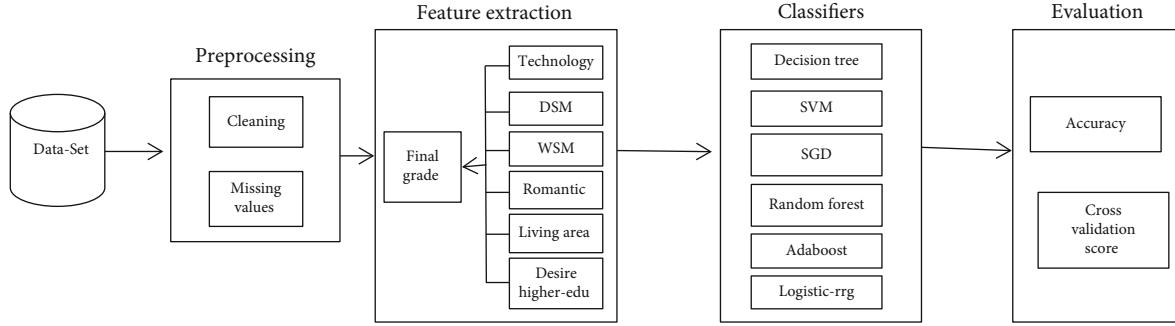


FIGURE 1: Educational data mining model steps.

by using machine learning algorithms. The dataset of 4000 students has attributes of online-duration, internet-traffic-volume, and connection-frequency, which were extracted, calculated, and normalized from the real internet usage. DT, NN, and SVM were used to find student educational performance by using these attributes. Internet-con and frequency attribute are positively linked, and internet-traffic-volume attribute is negatively linked with academic performance of students. The online-time and internet-time suffering results in surprising performance among different datasets. The number of features increases and improves accuracy. The results define that internet usage is able to distinguish and analyze student's academic performance.

Reference [13] narrated that in higher education, data mining approaches are used and create an attractive part in educational research. These approaches are used for identifying and finding meaningful data from large meaningless data. By using a supervised data mining method, find the results of student progress. To find the student progress is helpful for current educational organizations. The basic purpose of the study is to make a model with the help of classification methods. This model analyzes the student performance in Malaysia. This model is used to find the most important features from the large dataset. Many approaches which are KNN, naïve Bayes, DT, and logistic regression approaches are used to analyze the student academic result performance. These approaches are based on accuracy measure, precision, recall, and ROC curve. The output showing the naïve Bayes algorithm is better. NB is disclosing important attributes that are used to find excellent students whose grades are A+ and A.

Reference [14] reported that large number of students dropped out a major worry of higher education organizations. It greatly influences the fee of students and discarded public resources. It is necessary to find those students who are in danger of dropping out and find those attributes that are the cause of higher dropout rate. Educational data mining methods are used to recover this problem. In this study, the University Teknologi MARA students of computer science undergraduate after three years. DT, logistic regression, random forest, KNN, and NN algorithms are matched to analyze student performance. Several machine learning methods are combined and make an efficient model. The logistic regression method is the best algorithm to analyzing and predicting the dropout students.

3. Educational Data Mining Model

This study evaluates the impact of technology on student's educational performance. This study proposed the educational data mining (EDM) model that is divided into five major sections such as collection of dataset, preprocessing of dataset, feature extraction, selection of classifier, and model evaluation, see Figure 1. Each section may contain more than one subsection. In step one, dataset was cleaned and checked if there is no missing value. After cleaning the dataset, required features were extracted from the dataset like "use of technology," "weekly-social-media-use," or "weekday-social-media-use." In the next step, different learning models were used to predict student's final grade performance. After that, model's performance was compared based on accuracy score to select the best learner for the problem. The algorithms used in the study are DT, SVM classifier, SGD classifier, random forest classifier, AdaBoost, and logistic regression classifier.

3.1. Dataset Collection. The data used for the analysis is collected from an electronic-learning system called Kalboard 360 that is publicly available on <https://www.kaggle.com/d50stuck/kalboard-360-use-case>. The features of student dataset and their categories are listed in Table 1.

3.2. Preprocessing. In the preprocessing phase, first, we make sure that there is no irrelevant and unacceptable value existed inside the dataset. This process is called cleaning. After cleaning process, we analyzed the data and removed unnecessary fields that are not relevant to our research objective. This process makes data more refined and relevant to research objective. In the preprocessing, we also handled the null values in the dataset.

3.3. Selection of Classifier. After obtaining the required features, different classifiers were trained on the dataset. The algorithms used in the study are DT, SVM classifier, SGD classifier, random forest classifier, AdaBoost, and LR classifier.

3.4. Decision Tree. The DT classifier is simple and understandable by analysts and end users. It is a tree shape model built based on the features, see Figure 2. These are WSM, DSM, living area, romantic status, parent education, technology, and desire-higher-education; these all features are called nodes and influence the student final scores. Every

TABLE 1: Features of student dataset and their categories.

| Attribute | Description | Category of features |
|-------------------|------------------------------------------------|--------------------------------------------------------|
| Sex | Gender | |
| Age | Age of the student | |
| Address | Address of student | |
| Health | Health of the student | Demographical features |
| Living area | Urban and rural area | |
| Romantic status | Interest in romance | |
| Medu | Mother education | |
| Fedu | Father education | |
| Mother_job | Mother job | Participation of parents on the whole learning process |
| Father_job | Father job | |
| Study time | Study time | |
| Desire_higher_edu | Desire to receive higher education | |
| Final grade | Final grade of student (good, poor, fair) | |
| Period1_score | Subject 1 score | |
| Period2_score | Subject 2 score | Academic background features |
| Final-score | Final score of student on the basis of subject | |
| Freetime | Free time spend by student | |
| Absences | Absences of the students | |
| Failures | Failures of student | |
| Technology | Technology used by student | |
| DSM | Daily-social-media-usage | Technological features |
| WSM | Weekly-social-media-usage | |

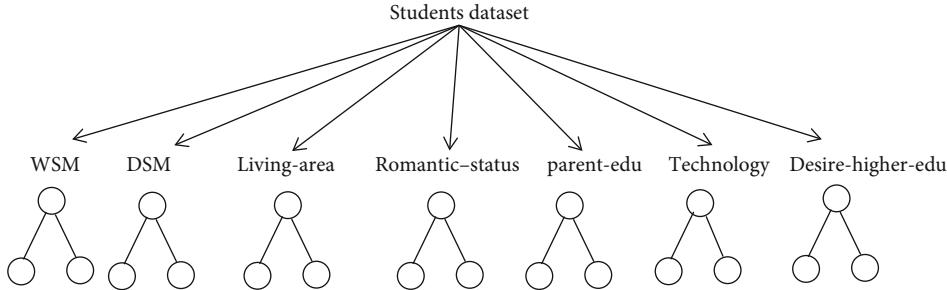


FIGURE 2: Decision tree classifier.

node is divided into subnodes. Every node makes decision on the basis of numeric value.

3.5. SVM Classifier. This classifier is a linear algorithm that is suitable for small datasets. Support vector machines are not suitable for large datasets because it takes small memory and needs more training time. I used this classifier because my dataset is small; it correctly classifies features that are influencing the student performance. It divides features into two classes, for example, living area divides into a rural and urban area and the use of technology divides the yes or no class. The study divides the use of internet into two classes such as “low use” and “high use”. If a person uses 1-2 days, then it is considered in “low use” and -1 weight is given to the user. If a person uses 4-5 days, then it is considered in “high use” and 1 weight is given to the user. In case of 3 days, 0 weight is given to it. All categories are shown in Figure 3.

X_1, X_2 , and X_n are vectors/features; y is a class to which features belong $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

$$\begin{aligned}
 W \cdot x + b &= 1, \\
 W \cdot x + b &= 0, \\
 W \cdot x + b &= -1.
 \end{aligned} \tag{1}$$

3.6. Random Forest. The whole tree is divided into small parts/samples. The random forest classifier builds small trees for every feature of the features like weekly-social-media-use, weekday-social-media-use, living area, romantic status, technology, and parent education. The random forest classifier analyzing the student performance by splitting the nodes of decision tree random forest builds multiple decision trees. At the end, voting process is performed for every sample and

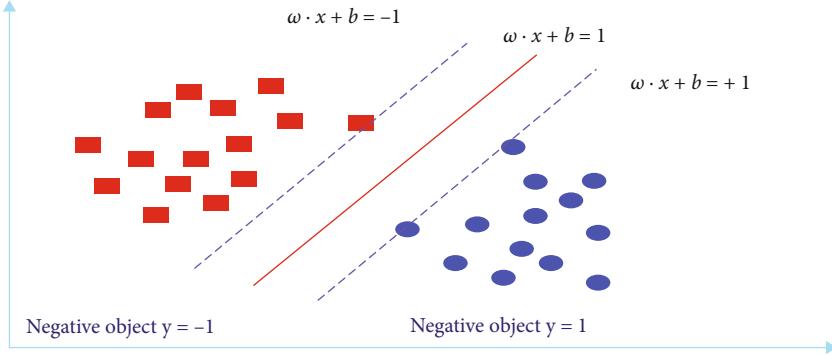


FIGURE 3: SVM classifier.

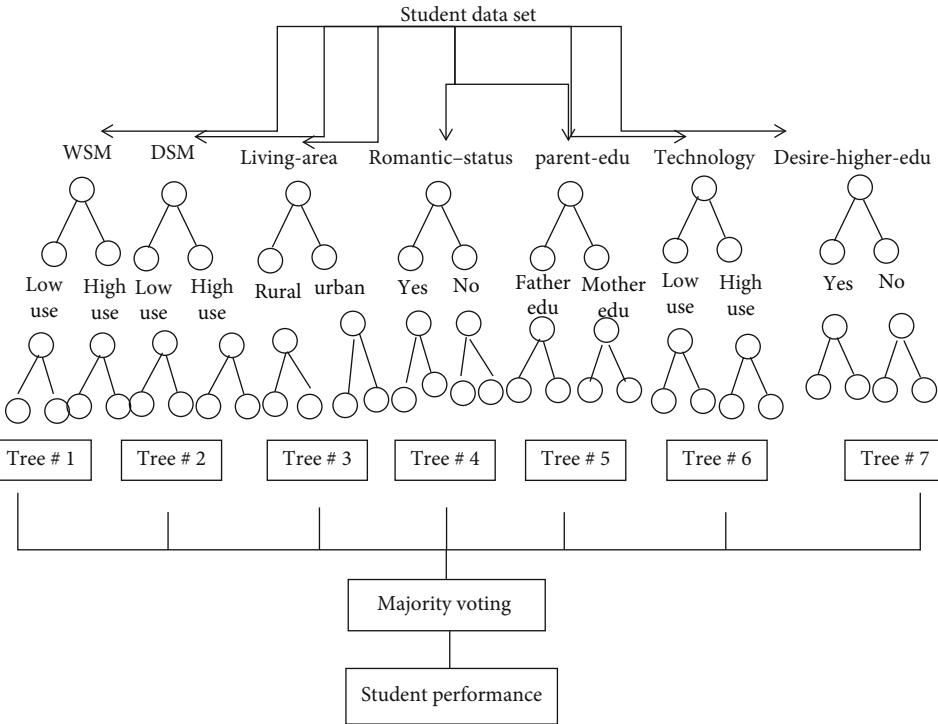


FIGURE 4: Random forest classifier.

finds the performance of the students. The random forest algorithm provides the best result then the decision tree algorithm and other algorithms that are used here. The working of the random forest classifier is as shown in Figure 4.

3.7. Logistic Regression Classifier. The logistic regression classifier is another type of supervised learning algorithm. It is also called the logit model. It is a statistical model used to find the probability of a class pass or fail. It uses a basic logistic function to construct a binary dependent variable. It is easy to implement. It provides a baseline to a binary classification. It also defines the link between dependent variable and independent variables. LR outcome is constant. The equation of logistic regression is represented in

$$y = \frac{e^{b_0 + b_1 * x}}{1 + e^{b_0 + b_1 * x}}, \quad (2)$$

where y is the predictive output, b_0 is the intercept term, and b_1 is the single value coefficient of input (x).

3.8. AdaBoost. The AdaBoost classifier is a meta-algorithm of machine learning. Meta-algorithms mean different low accuracy classifiers merged into a single highly predictive model to increase performance. This classifier is sensitive to error porn data and outliers. This algorithm is less risky in overfitting problems as compared to other algorithms. The AdaBoost classifier is used to build a high-performance classifier whose accuracy is high. This classifier combines weak and poor classifiers and makes a strong and highly performing classifier.

As shown in Figure 5, the AdaBoost classifier works in the following steps:

- (1) Firstly, AdaBoost selects training samples randomly

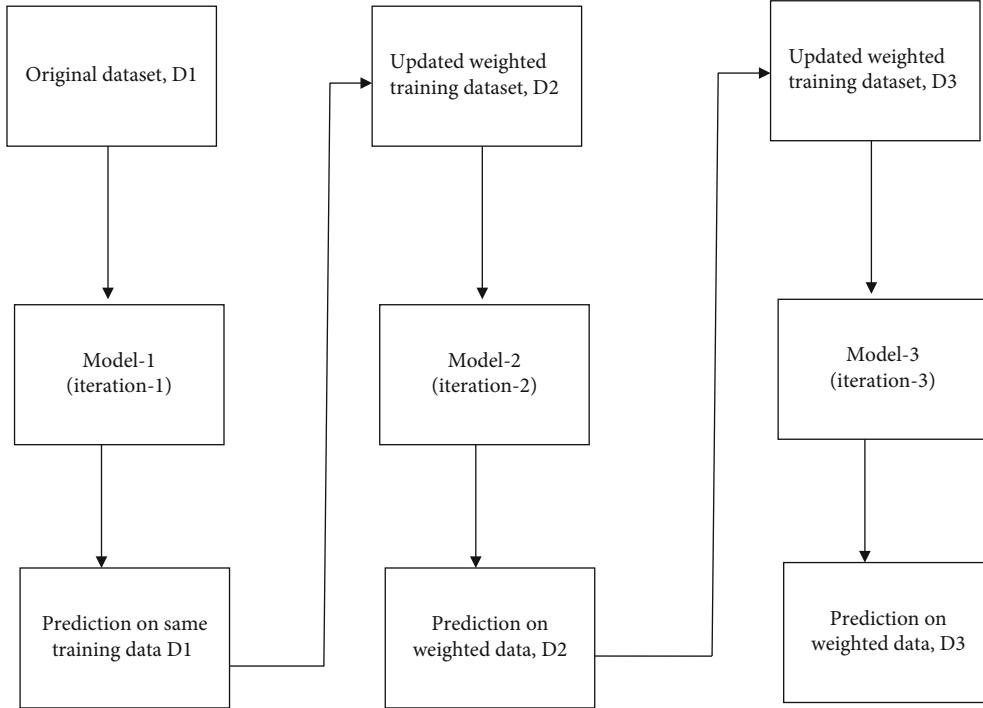


FIGURE 5: AdaBoost classifier.

- (2) It trains the AdaBoost machine learning algorithm by selecting the samples based on the correct analysis of the last training
- (3) It allocates the higher weight to wrong classified samples so the next repetition of classification gets the high probability for classification
- (4) It allocates the weight to the trained classifier in each repetition according to the accuracy of the classifier. It generates a high-performance classifier
- (5) This process repeats until the complete training samples fits without any error
- (6) To perform voting process on all the learning algorithms you generate

3.9. SGD Classifier. Stochastic Gradient Descent (SGD) is an optimization technique. In the Stochastic Gradient Descent approach, complete dataset is not selected; some samples are selected randomly. Total samples are called a batch. The batch is created from the complete dataset. The complexity is high when the dataset is big. It uses a single sample of data. The next time the sample is exchanged with the next sample randomly and then performs repetition. It increases the efficiency of the classifier and is easy to implement.

4. Results

The results of the educational data mining model are presented as follows.

| | | Predicted class | |
|--------------|---|-------------------------|-------------------------|
| | | N | P |
| Actual class | P | True positive (TP) | False negatives (FN) |
| | N | False positives (FP) | True negatives (TN) |

FIGURE 6: Confusion matrix [15].

4.1. Model Evaluation. The evaluation phase of our model analyzes the outcomes of every classifier on the basis of the following factors. Confusion or error matrix is used to evaluate the performance of a classifier, see Figure 6.

Accuracy is the basic evaluation metric to analyze the rate of correctness of the prediction. The accuracy is measured with a formula, see the following:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}. \quad (3)$$

4.2. Correlation Heat Map. A heat map is a simple and useful tool to find out useful attributes in a dataset. Diagram represents correlation between different features. Value 1 shows two feathers are positively correlated, and when the correlation is closer to or similar to -1 increase or decrease, one variable value will decrease or increase the other variable. The main advantage to use a heat map is how a feature is useful according to my problem and cleans my dataset before its use and execution. The correlation heat map is shown in Figure 7.

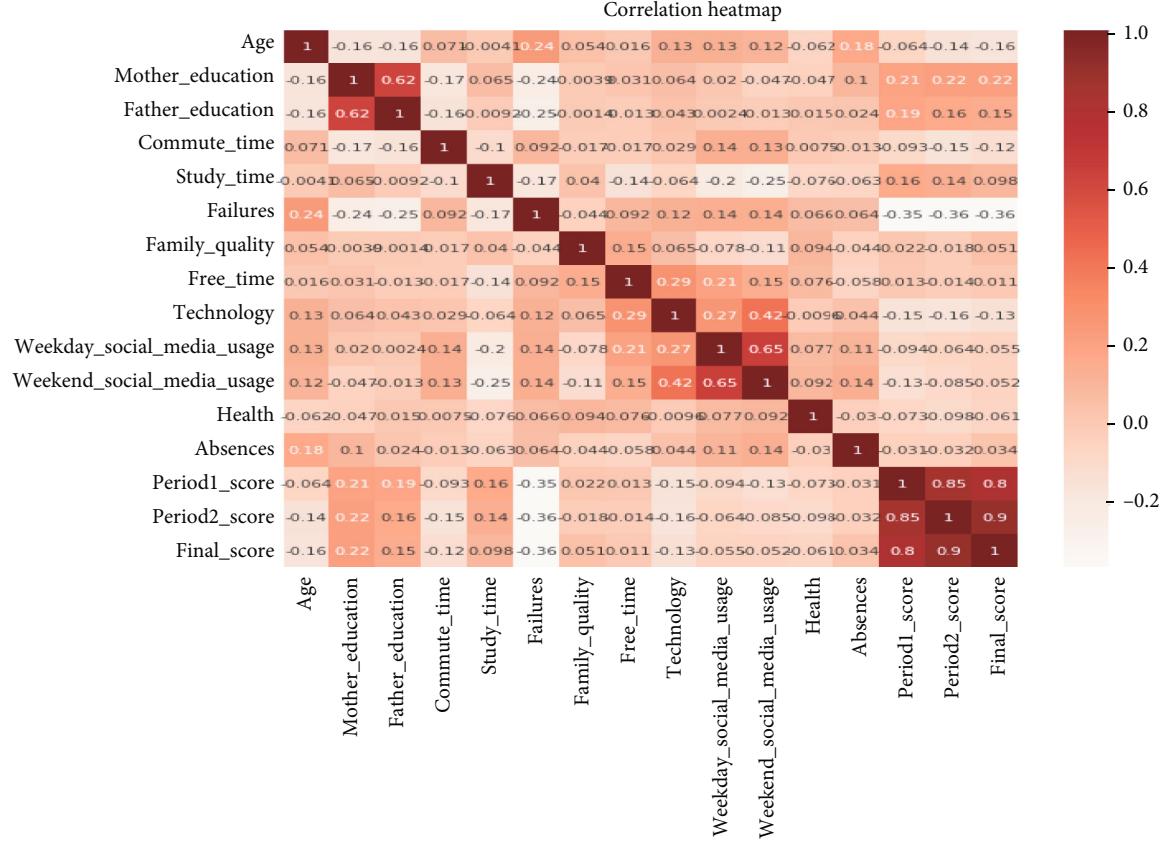


FIGURE 7: Correlation heat map.

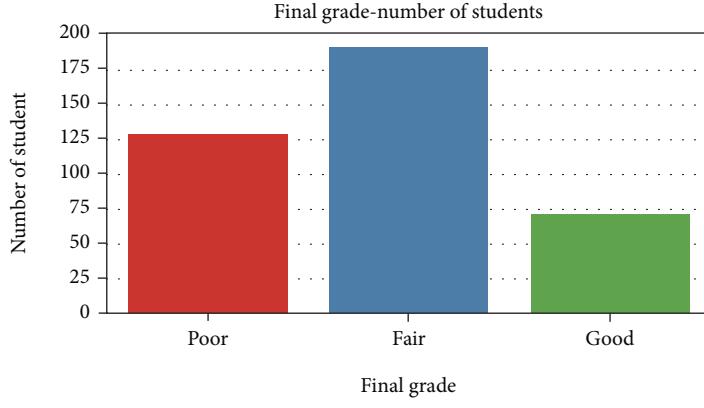


FIGURE 8: Final grade distribution.

4.3. Final Grade Distribution. The author classifies students into three categories, “good,” “fair,” and “poor” according to their final exam performance, and then analyzed a few features that have a significant influence on students’ final performance, including romantic status, parent education level, frequency of going out, desire of higher education, and living area. The three categories of students are shown in Figure 8.

4.4. Final Grade by Frequency of Technology Usage. Figure 9 shows 5 levels of technology understanding and use. The use of technology depends on the understanding of devices that

are used by students. Good students understand the technology and use it for their studies. The performance is high because they understand the technology and cannot waste their time. The poor students are not capable of using technology. When they use devices, they cannot understand what they are working, so they waste their time and energy. So, that type of student has low performance. The fair students are that some students understand, and some cannot understand the technology; some students use technology but cannot improve their performance because they have no proper guidelines and training to use technology.

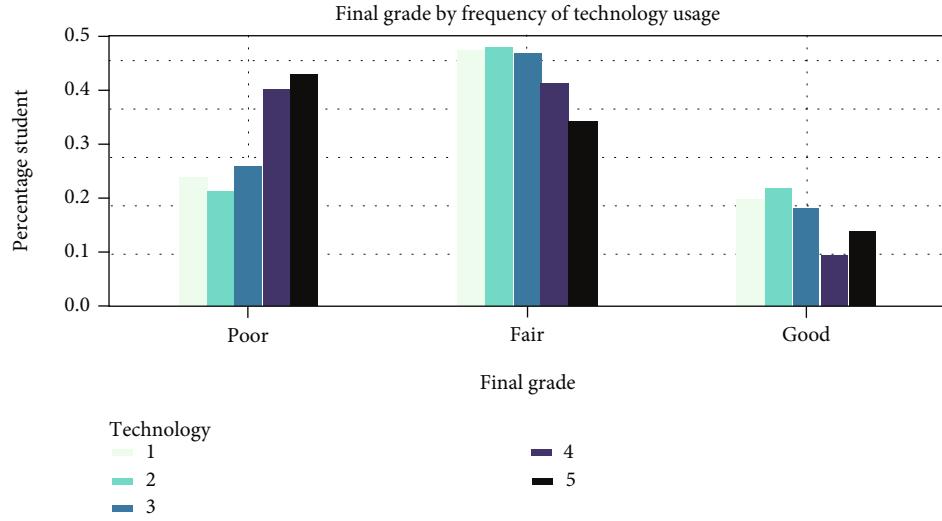


FIGURE 9: Final grade by frequency of technology usage.

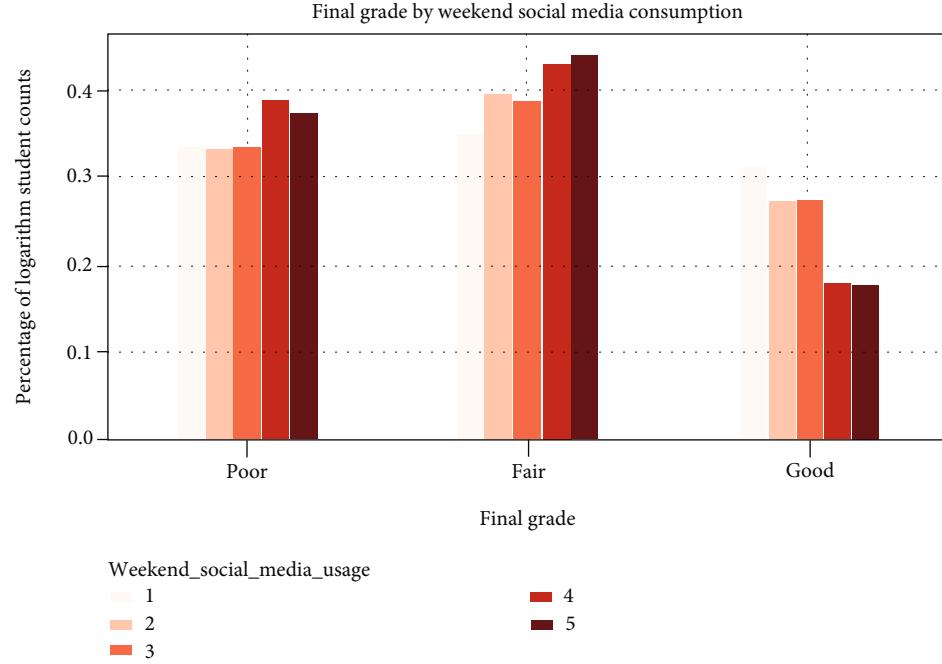


FIGURE 10: Weekend-social-media consumption with final grade.

4.5. Final Grade by Social Media Consumption. Figure 10 shows good or intelligent, fair or normal, and poor or weak students use weekly social media. The use of social media divides into five levels (1, 2, 3, 4, and 5). Low social media usage on the weekend is represented by 1 and 2. The highest use of social media is represented by 4 and 5 levels. The medium use of social media on the weekend is represented by 3. Poor students that are weak in their studies use social media which results in their performance becoming slower. Fair and good students use social media highly; then, their performance is also decreasing. So, the high use of social media also influences the student performance.

4.6. Final Grade by Parents' Education Level. Parents' education level influences student performance.

4.6.1. Father Education Level. A parent's education level has a positive correlation with a student's final score. Father education level influences the student grade. Educated fathers also affect children's education. Father education affects the student's performance, but mother education greatly affected the student performance. Figure 11 shows that father education influences the student grade.

4.6.2. Mother Education Level. Comparatively, the mother's education level has a bigger influence than the father's

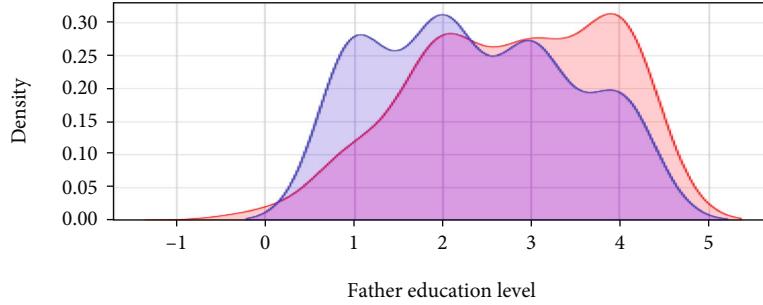


FIGURE 11: Father education level.

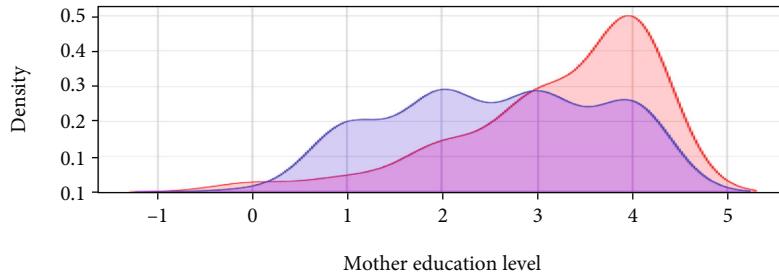


FIGURE 12: Mother education level.

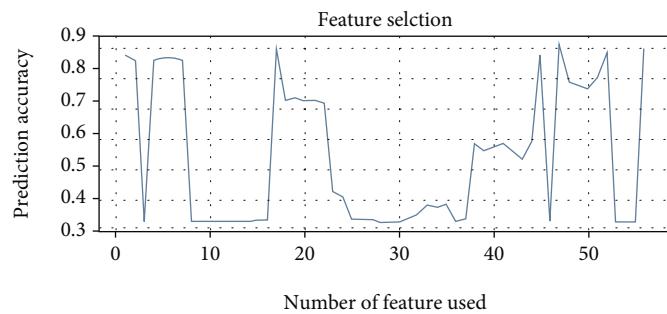


FIGURE 13: Feature selection.

education level. Because the mother guides their children and supports them in their studies more than fathers so mother education highly influences student final score. Most mothers are uneducated; the student performance in early stage is based on mother education. Some mothers are non-serious about the studies of their child because they are not educated, so their child cannot perform well in education at school level. Figure 12 shows how much mother education influences the student performance.

4.7. Feature Effect. Figure 13 shows that the number of features increases the prediction accuracy of a classifier. Multiple features help the classifier to train and get accurate results. But condition is that features are correlated to the problem. Relevant features greatly influence the accuracy, but irrelevant features decrease the accuracy. On the other hand, multiple features can complex the classifiers and some classifier like SVM cannot work on large number of features because it has a limited memory. Deep learning correctly classifies the large number of features. So, we can say that relevant large number of features improves the accuracy,

but the multiple features also increase the complexity of a classifier. Sometimes, using multiple features cannot increase the prediction accuracy because the features are irrelevant to the problem, and sometimes, a small number of features greatly influence the prediction accuracy.

4.8. Final Results of Classifiers. Supervised learning algorithms used to predict the student academic performance with the use of technology. These algorithms are DT, random forest, SVM, L-regression, AdaBoost, and SGD. The score of decision tree is 0.90%, random forest score is 0.98%, support vector classifier score is 0.86%, logistic regression is 0.88%, AdaBoost is 0.89%, and SGD classifier score is 0.84%. The scores prove that the random forest classifier has the best results as compared to other classifiers. The DT classifier is the second one. The decision tree classifier score is lower than random forest because decision tree has problem of overfitting. The Stochastic Gradient Descent classifier gains the lowest scores. The comparison of classifiers is shown in Table 2.

TABLE 2: Classifier results.

| Sr. | Classifier | Score | Cross-validation score |
|-----|---------------------|-------|------------------------|
| 1 | DT | 0.90 | 0.87 |
| 2 | Random forest | 0.98 | 0.86 |
| 3 | SVM | 0.86 | 0.87 |
| 4 | Logistic regression | 0.88 | 0.31 |
| 5 | AdaBoost | 0.89 | 0.91 |
| 6 | SGD | 0.84 | 0.77 |

5. Discussion

Last few years' educational data mining received great attention. Many data mining approaches extract knowledge from educational databases. The extracted information from educational data helps the educational institutes to improve teaching and learning process. This enhancement of data improves the students and educational institute output performance. Student behavioral attributes also influence the student performance. Using behavioral feature accuracy of classifiers is greater than without using behavioral features. The decision tree classifier shows higher accuracy of 75% without behavioral feature of 55% percent accuracy. The newly emerging field of research is EDM. Educational DM is the combination of data mining and educational data [15–18]. It helps the student to improve their performance and their learning activities. The educational data is used from any education repository, for example, learning management system, web-based education, and online data. Assembling methods are also used for getting higher performance of a classifier. These techniques divide the data into equal sizes, and the voting process is used. Highly voted data is extracted and concluded the results. The bagging algorithms are bagging, boosting, random forest, etc. In traditional models, single model is used for training data, however, in ensemble models, more than one model is used for training of data. Multiple models train with attribute with voting process. The advantage of the assembling method is that the accuracy is higher than the single model [19–22].

Determining assessment and activity data can affect students' educational performance. The four selection algorithms are decision tree, random forest, multilayer perceptron, and logistic regression which were used to identify the important features that affect students' academic performance. Results show that the most important feature that can affect student educational data is assessment data like final exam and assignment marks are most important. Decision tree performs useful as using random forest achieving highest accuracy [23–26].

Technology like social media usage (weekday and weekly), living area, parent education, desire to receive higher education, and romantic status are features that greatly influence the student performance. Social media use natively and consuming more time on social media can decrease the student performance. By using the random forest assembling method, the accuracy of a classifier is higher 98% than other classifiers like decision tree, AdaBoost, logistic regression, and Stochastic Gradient Descent. The tech-

nology feature greatly influences the student performance [27]. Behavioral, assessment marks, parent involvement, living area, and many more features can influence the student education, but in this modern world, technology can play a vital role in every student's life. This paper focuses on technology like social media uses. The main benefit of assembling an algorithm is gaining higher accuracy than the single model classifier like the SVM classifier.

6. Conclusion

Academic achievement is the biggest concern of every educational institute. This paper describes the importance and impact of technology on student education. This study used machine learning techniques such as DT, SVM classifier, SGD classifier, random forest classifier, AdaBoost, and logistic regression classifier to determine the patterns inside the student performance data. We have used six different classifiers to analyze the student performance records. Our objective was to evaluate and analyze the impact of technology on student education, so we use our attributes including technology features, weekday-social-media-use, and weekend-social-media-use to analyze student data. All six of our classifiers achieved performance by adding technology features along with other features. Notably, random forest achieved higher accuracy of a classifier which is 98% as compared to the other classifiers. The score of decision tree, AdaBoost, logistic regression, SVM, and SGD is 90%, 89%, 88%, 86%, and 84%, respectively. This research shows that nowadays technology is a very important factor to achieving better performance in educational institutes. Social media greatly influence the student education. This feature increases the accuracy of classifiers. Currently, in the changing world, online and home-based educations become very important. Furthermore, the analysis of the different factors of the technology, the negative impact of technology, and the impact of home-based learning could be the key direction of future research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

We appreciate the partial collaboration with Universiti Malaysia Sabah, Malaysia, in this work.

References

- [1] M. S. Chen, J. Han, and P. S. Yu, "Data mining: an overview from a database perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, pp. 866–883, 1996.

- [2] V. Ramesh, P. Parkaviand, and K. Ramar, "Predicting student performance: a statistical and data mining R. Sivakumar, "Effects of Social Media on Academic Performance of the Students", "The Online Journal of Distance Education and e-Learning, vol. 8, no. 2, p. 90, 2020.
- [3] S. Naseem, A. Alhudhaif, M. Anwar, K. N. Qureshi, and G. Jeon, "Artificial general intelligence based rational behavior detection using cognitive correlates for tracking online harms," Personal and Ubiquitous Computing, 2022.
- [4] J. Cradler, M. McNabb, M. Freeman, and R. Burchett, "How does technology influence student learning," Learning and Leading with Technology, vol. 29, no. 8, pp. 46–49, 2002.
- [5] P. Sokkheyand and T. Okazaki, "Developing web-based support systems for predicting poor-performing students using educational data mining techniques," Studies, vol. 11, no. 7, 2020.
- [6] A. Sadiq, M. Anwar, R. A. Butt et al., "A review of phishing attacks and countermeasures for internet of things-based smart business applications in industry 4.0," Human Behavior and Emerging Technologies, vol. 3, no. 5, pp. 854–864, 2021.
- [7] V. P. Breslelean, "Analysis and predictions on students' behavior using decision trees in Weka environment," in 2007 29th International Conference on Information Technology Interfaces, pp. 51–56, Cavtat, Croatia, 2007.
- [8] M. Anwar, F. Masud, R. A. Butt, S. M. Idrus, M. N. Ahmad, and M. Y. Bajuri, "Traffic priority-aware medical data dissemination scheme for IoT based WBASN healthcare applications," Computers, Materials & Continua, vol. 71, no. 3, pp. 4443–4456, 2022.
- [9] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k means clustering algorithm for prediction of students academic performance," 2010, <https://arxiv.org/abs/1002.2425>.
- [10] S. K. Yadav and S. Pal, "Data mining: a prediction for performance improvement of engineering students using classification," 2012, <https://arxiv.org/abs/1203.3832>.
- [11] K. N. Qureshi, E. Ahmad, M. Anwar, K. Z. Ghafoor, and G. Jeon, "Network functions virtualization for mobile core and heterogeneous cellular networks," Wireless Personal Communications, vol. 122, no. 3, pp. 2543–2559, 2022.
- [12] T. M. Lakshmi, A. Martin, R. M. Begum, and V. P. Venkatesan, "An analysis on performance of decision tree algorithms using student's qualitative data," International Journal of Modern Education and Computer Science, vol. 5, no. 5, pp. 18–27, 2013.
- [13] M. Faheem, S. B. H. Shah, R. A. Butt et al., "Smart grid communication and information technologies in the perspective of Industry 4.0: opportunities and challenges," Computer Science Review, vol. 30, pp. 1–30, 2018.
- [14] R. A. Butt, M. Faheem, M. Anwar, K. H. Mohammadani, and S. M. Idrus, "Traffic aware cyclic sleep based power consumption model for a passive optical network," International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, 2022.
- [15] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm," Procedia Technology, vol. 25, pp. 326–332, 2016.
- [16] M. Mohammadi, M. Dawodi, W. Tomohisaand, and N. Ahmadi, "Comparative study of supervised learning algorithms for student performance prediction," in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIC), pp. 124–127, Okinawa, Japan, 2019.
- [17] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," Computers in Human Behavior, vol. 98, pp. 166–173, 2019.
- [18] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacoband, 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.
- [19] W. W. Yaacob, N. M. Sobri, S. M. Nasir, N. D. Norshahidi, and W. W. Husin, "Predicting student drop-out in higher institution using data mining techniques," Journal of Physics: Conference Series, vol. 1496, p. 012005, 2020.
- [20] M. Anwar, A. H. Abdullah, A. Altameem et al., "Green communication for wireless body area networks: energy aware link efficient routing approach," Sensors, vol. 18, no. 10, p. 3237, 2018.
- [21] S. A. Salloum, M. Alshurideh, A. Elnagarand, and K. Shaalan, "Mining in educational data: review and future directions," in Joint European-US Workshop on Applications of Invariance in Computer Vision, pp. 92–102, Springer, Cham, 2020.
- [22] E. A. Amrieh, T. Hamtinian, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," International Journal of Database Theory and Application, vol. 9, no. 8, pp. 119–136, 2016.
- [23] M. Anwar, A. H. Abdullah, R. A. Butt, M. W. Ashraf, K. N. Qureshi, and F. Ullah, "Securing data communication in wireless body area networks using digital signatures," Technical Journal, vol. 23, no. 2, pp. 50–55, 2018.
- [24] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantisand, and O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," Applied Sciences, vol. 10, no. 1, p. 90, 2020.
- [25] M. Anwar, A. H. Abdullah, K. N. Qureshi, and A. H. Majid, "Wireless body area networks for healthcare applications: an overview," Telkomnika, vol. 15, no. 3, pp. 1088–1095, 2017.
- [26] M. Anwar, A. H. Abdullah, R. R. Saedudin, F. Masud, and F. Ullah, "CAMP: congestion avoidance and mitigation protocol for wireless body area networks," International Journal of Integrated Engineering, vol. 10, no. 6, pp. 59–65, 2018.
- [27] A. Alhassan, B. Zafar, and A. Mueen, "Predict students' academic performance based on their assessment grades and online activity data," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 4, 2020.