Machine learning based COVID-19 study performance prediction

Md. Ataur Rahman[1], Md. Sadekur Rahman[2], Mohammad Monirul Islam[2], Mahady Hasan[1,3], Md. Tarek Habib[1,3]

[1]Fab Lab IUB, Independent University, Bangladesh, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[3]Department of Computer Science and Engineering, Independent University, Bangladesh, Dhaka, Bangladesh

Article Info

Article history:

Received Dec 14, 2023
Revised Nov 5, 2024
Accepted Nov 14, 2024

Keywords:

Accuracy
Machine learning
Performance metrics
Prediction system
Student performance

ABSTRACT

COVID-19 has impacted education worldwide. In this troublesome situation, it is hard enough for an institution to predict a student's performance. Students' performance prediction has always been a complex task and this pandemic situation has led this task to be more complex. The main focus of this work is to come up with a machine learning model based on a classical machine learning technique to predict the change in students' performance due to COVID-19. Initially, some relevant features are selected, based on which the data are collected from students of some private universities in Bangladesh. After the entire data set is formed, we preprocessed the dataset to remove redundancy and noise. These preprocessed data are used for testing and training using the proposed model. The model is extensively evaluated in this way using three separate classical machine learning techniques, namely linear regression, k-nearest neighbors (k-NN), and decision tree. Finally, the results of the entire experiment follow, demonstrating the power of the machine learning model in such an application. It is observed that the proposed model with linear regression exhibits the best performance with an R^2 error of 0.07% and an accuracy of 99.84%.

Corresponding Author:

Md. Ataur Rahman
Fab Lab IUB, Independent University, Bangladesh
Dhaka, Bangladesh
Email: ataurrahman862828@gmail.com

1. INTRODUCTION

Education is one of the basic elements in a country's development. In recent years, the value of higher education has increased. Students' performance has a big impact on the learning process. Predicting student performance is key to identifying students through which future outcomes are likely. This information may be helpful and used in predictions if the data is processed in a knowledgeable method.

A major area of study is called educational data mining (EDM), which utilizes a variety of algorithms to improve educational results and clarify educational processes for future decision-making. EDM is the process of converting unusable educational system data into data that is helpful for decision-making based on data. Much research has been done in the field of data mining in education to develop predictive models for predicting student performance and identifying at-risk students [1].

Global education has been impacted by COVID-19. It is challenging enough for a school to forecast a student's success in this problematic situation. Predicting student performance has always been difficult, and the pandemic situation has made it even more difficult. Thus, a typical application problem of EDM is to predict the change in a student's performance due to COVID-19 or a similar pandemic. To solve the problem, we propose a machine learning model based on a classical machine learning technique to predict the change in a student's performance due to the COVID-19 pandemic.

This system can benefit in some ways. Educational institutions can use the solution application to the problem to estimate and comprehend the future performance of their students affected by not only COVID-19 but also any pandemic so that they can take precautionary measures to mitigate loss and/or damage. Students can use this application to assess their performance degradation due to the hindrance of a pandemic so that they can take adaptive action to handle the situation as efficiently as possible. This application can also be used by the guardians of the students to assess their performance due to the hindrance of any pandemic so that they can be alert in advance.

We study relevant papers from various national and international journals, as well as parts from periodicals, newspapers, websites, and conference proceedings. After that, we speak with students to learn about the benefits and drawbacks of studying during COVID-19. Thus, eleven features have been selected for the proposed model to solve the problem. We collect thousand and nineteen (1,019) data from private university students in Bangladesh to build the entire data set. In the prepared data set, there are many unnecessary data as well as missing data. Then we preprocess the data set by using two operations filling in the missing value and noise removal. We employed three separate classical machine learning algorithms, namely linear regression, decision tree, and k-nearest neighbors (k-NN) to build the proposed model. We perform extensive and in-depth modeling separately for the three separate machine learning algorithms. The model with linear regression produces the best results.

In this paper, machine learning modeling is done using off-the-rack machine learning algorithms to predict student performance due to a pandemic like COVID-19, which extends the state-of-the-art knowledge in solving a novel real-life problem. In essence, the principal contributions of this research work to the current state of literature are:

- A demonstration of the deployment of a machine learning model based on state-of-the-art regression techniques to solve a novel real-life problem.
- The first attempt to solve the problem of the prediction of the change in study performance due to a pandemic like COVID-19.
- Systematically formation of a strong base of distinguishing features for regression to successfully predict the change in study performance due to a pandemic like COVID-19.
- An in-depth exploration of different regression algorithms in the context of the prediction of change in study performance to facilitate future research.

The remaining sections of the paper are arranged as follows: section 2 describes the recent state of solutions to address the problem of students' performance prediction and its variants. Section 3 describes the proposed model and the architecture of the proposed prediction system. Section 4 discusses the entire research methodology. Section 5 describes the experimentation procedure in detail, which is followed by the analysis of results by comparing them with other works in section 6. Finally, a summary, limitations, and future scope are presented in section 7.

2. RELATED WORKS

There have been some noticeable research works on student performance prediction as well as its variations. Some work has been done on the prediction of student performance irrespective of COVID-19 or any other pandemic. Again, some predictive research works have been performed taking the COVID-19 pandemic into account.

Tarik *et al*. [2] encourage the development of particular talents. By doing both individual and group interviews with students, the educational guidance counselor builds the educational support service. Mandayam *et al*. [3] utilized the COVID-19 time-series dataset and two supervised learning models to provide future predictions. To evaluate the effectiveness of prediction, linear regression, and support vector regression are compared. Prakash *et al*. [4] found that artificial intelligence is the most effective weapon against the COVID-19 issue. Two of the key technologies in artificial intelligence are machine learning and deep learning. Rai *et al*. [5] found that understanding levels and giving students the learning they require through machine learning were employed. To get the best prediction results, the random forest algorithm is applied. Hossain *et al*. [6] describe the important initiatives that the government could implement to further develop online education and distance learning (DL) in Bangladesh. In 2006, during a time of significant changes to Bangladesh's infrastructure, this study was conducted on the impact prediction of online education. Sultana and Khan [7] discuss how satisfied students are with their online education. Tabassum *et al*. [8] provide information on the situation involving educators and online learning facilities. Online education has been proven to be related to happiness and residence areas. Few studies have also been done from the perspective of the instructor. According to Rybarczyk and Zalakeviciute [9], there are 24 models built using machine learning that predict the air quality as if the COVID-19 quarantine had never existed and are based on the gradient boosting machine algorithm. This research suggests a different strategy that involves teaching a machine learning system to recognize how time and weather affect air pollution. In regions with difficult terrain, it also exhibits decreased performance. Decreases in particular matter $PM_{2.5}$ concentrations vary similarly depending on the kind of monitoring location. Acharya and Sinha [10] used decision trees and other machine-learning approaches to predict students' performance. Vandamme *et al*. [11] employed machine learning methods such as decision trees, linear discriminate analysis, and neural networks to predict students' early performance. Cortez and Silva [12] four classification approaches were used to improve student performance and help achieve the goal: decision trees, random forests, artificial neural network (ANN), and support vector machines (SVM). Mia *et al*. [13] using a variety of machine learning techniques, it is possible to predict how many students will enroll in a semester at a private university. They showed that SVM had the best accuracy.

Machine learning techniques are utilized in fields other than education. Machine learning techniques are also applied in the medical field and others. Ferdowsy *et al*. [14] worked on the risk of obesity predicted using machine learning methods. Their paper's main argument is that individuals should be aware of their risk of obesity and the causes of their obesity. Arif *et al*. [15] also worked on the risk of becoming drug or alcohol addicted, which was predicted using a variety of machine-learning approaches. To determine the risk of drug addiction, they employed nine machine learning algorithms. Rimi *et al*. [16] employed a machine learning technique to forecast dental disease based on how a nation's citizens behave regularly. Chen *et al*. [17] presented a 94.8% accurate machine learning method for predicting chronic disease outbreaks using hospital data and a convolutional neural network (CNN)-based multimodal algorithm. Xu *et al*. [18] described a unique machine learning strategy for predicting student performance in degree programs that addresses background diversity, course informativeness, and dynamic progress, with improved results on a University of California dataset.

Abdullah *et al*. [19] using variables including make, model, age, mileage, and location, random forest machine learning predicts Bangladesh's reconditioned car pricing with 99.59% accuracy. It then evaluates mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE), and R^2, supporting the decisions of stakeholders. Barhmi and El Fatni [20] present four hybrid models for wind speed forecasting that combine neural networks with SVM. The neural models perform better than the SVM models, as seen by their reduced MSE, mean error (ME), and MAE values. Abdullah *et al*. [21] analysis of the effects of COVID-19 e-learning on students at Jordan University of Science and Technology (JUST) shows that machine learning regression predicts academic success. Evaluation reveals the lowest MAE and RMSE for every course. Grari *et al*. [22] evaluated them using the R^2 score, MAE, and MSE to assess their predictive performance.

Considering the discussion above meticulously, we can come up with the conclusion that not a single significant research work has been performed on the prediction of student performance owing to the COVID-19 pandemic. This reveals a clear research gap in this respect. This is quite interesting for researchers in EDM.

3. PROPOSED MODEL

Initially, it is assumed that a user must answer a few questions on student performance. Through the web-based application, users may quickly provide answers to questions. The user's information will be forwarded to the server, and subsequently to the expert system. Data will be collected via the expert system interface after this expert system has been processed. The obtained data will be connected to the previously built forecasting model by the interface engine and transmitted to the collected knowledge-based system. After the prediction process is completed, it will return the result to the expert interface. Based on the input provided, the results will be generated by applying the proper methods to the refined data. The user will then get the outcomes via the server, providing insight into the probability of study performance changes due to COVID-19. The entire process for predicting study performance change due to COVID-19 is depicted in Figure 1 in terms of system architecture.

3.1. Description of features

The data was collected from the students through various questions. These data came from students currently studying in different semesters. Tables 1 and 2 show the attributes, i.e. features of the data sets along with some relevant information. Table 1 shows the features utilized without the COVID-19 pandemic, whereas Table 2 shows the features used only during the COVID-19 pandemic. So, Table 1 represents the features in the general situation and Table 2 represents the demographic information in the form of features that come into play during COVID-19. The features in general, as shown in Table 1, are semester no., total credits, theory subject credits, lab subject credits, non-departmental subject credits, cumulative grade point average (CGPA), and semester grade point average (SGPA), where SGPA is the output variable and all other features are input variables. Likewise, the demographic features are financial problems due to the COVID-19 situation, the extent of meeting friends during the COVID-19 situation, sleeping time in the unit of hours during the COVID-19 situation, sleeping time in hours before the COVID-19 situation, time spent in the unit of hours on social media during the COVID-19 situation, and SGPA, where SGPA is the output variable and all other features are input variables. Thus, the output feature without COVID-19 ($y_{\bar{C}}$) and with COVID-19 (y_C) becomes as follows:

$$y_{\bar{C}} = (x1, x2, x3, \ldots, x6) \quad (1)$$

$$y_C = (x1, x2, x3, \ldots, x11) \quad (2)$$

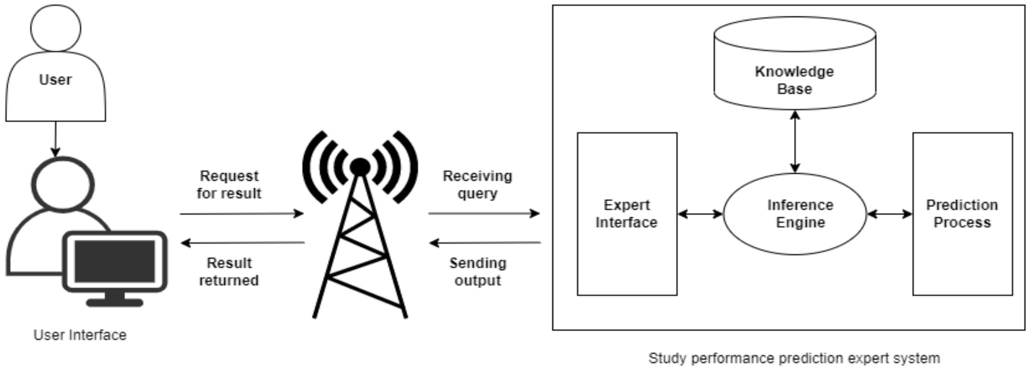

Figure 1. The system architecture of the change in the study performance prediction system due to COVID-19

Table 1. The attributes used in general, i.e. without the COVID-19 situation

Feature name	Representation	Feature type	Type of value	Range
Semester no.	x_1	Input	Integer, i.e. discrete	1-12
Total credit	x_2	Input	Integer, i.e. discrete	1-25
Theory subject credit	x_3	Input	Integer, i.e. discrete	1-20
Lab subject credit	x_4	Input	Integer, i.e. discrete	1-12
Non-departmental subject credit	x_5	Input	Integer, i.e. discrete	1-15
CGPA	x_6	Input	Real	[0, 4]
SGPA	$y_{\bar{C}}$	Output	Real	[0, 4]

Table 2. The demographic attributes used in the COVID-19 situation

Feature name	Representation	Feature type	Type of value	Range
Financial problems	x_7	Input	Yes/NO	Binary, i.e. {0, 1}
The extent of meeting friends	x_8	Input	Very Low/Low/ Medium/High	{0, 1, 2, 3}
Sleeping time in hours	x_9	Input	Real	[0, 16]
Sleeping time before the COVID-19 pandemic	x_{10}	Input	Real	[0, 16]
Time spent in hours on social media	x_{11}	Input	Real	[0, 16]
SGPA	y_C	Output	Real	[0, 4]

3.2. Performance measures

Regression algorithms predict output values based on input characteristics from data provided to the system. Regression algorithms utilize the characteristics of the training data to generate a model. That analysts then use to forecast new values.

The actual prediction process involves feeding new data to the trained model and using the model's algorithms to generate predictions. In this case, two actual predictions have been extracted. One with COVID-19 actual prediction metric (A_C) and the other without COVID-19 actual prediction metric ($A_{\bar{C}}$). Then the difference between with COVID-19 actual prediction metric (A_C) and without COVID-19 actual prediction metric ($A_{\bar{C}}$) is represented by *AP* and calculated in the following way:

$$AP = A_C - A_{\bar{C}} \quad (3)$$

In this case, input all the test data into the model and get the output. After that, all the outputs are sums. Here, two model predictions have been extracted. One with COVID-19 model prediction metric (M_C) and the other without the COVID-19 model prediction metric ($M_{\bar{C}}$). Then the difference between with COVID-19 model prediction metric (M_C) and without the COVID-19 model prediction metric ($M_{\bar{C}}$) is represented by *MP* and calculated in the following way:

$$MP = M_C - M_{\bar{C}} \tag{4}$$

Finally, the intended performance metric (*PM*) is calculated as shown in (5).

$$PM = \begin{cases} \left(1 - \frac{\sum_{e=1}^{D_{test}} AP - MP}{|D_{test}|}\right) \times 100\% & if\ PM\ is\ accuracy \\ \left(\frac{\sum_{e=1}^{D_{test}} AP - MP}{|D_{test}|}\right) \times 100\% & if\ PM\ is\ error \end{cases}, \tag{5}$$

where *PM* is one of the five metrics accuracy, MAE, R^2, MSE, and RMSE.

4. RESEARCH METHODOLOGY

These data came from students currently studying in different semesters. The next part will go through data collection and pre-processing procedures in detail. The data collected was in text format and therefore unsuitable for training. The data is subsequently presented in a machine learning algorithm using the data cleaning approach. In three phases, we employed three machine-learning algorithms. The accuracy was first calculated on the processed data before the principal component analysis (PCA), then again after the PCA, and finally on the unrefined data using the algorithm. We have analyzed the classifiers using accuracy as well as other metrics, including accuracy, MAE, R^2, MSE, and RMSE. These work processes are shown in the flow diagram in Figure 2. The method for getting the process data set is shown in Figure 2, starting with data collection on the left. On the other side, the three machine learning algorithms utilized for decision making are listed at the bottom right.

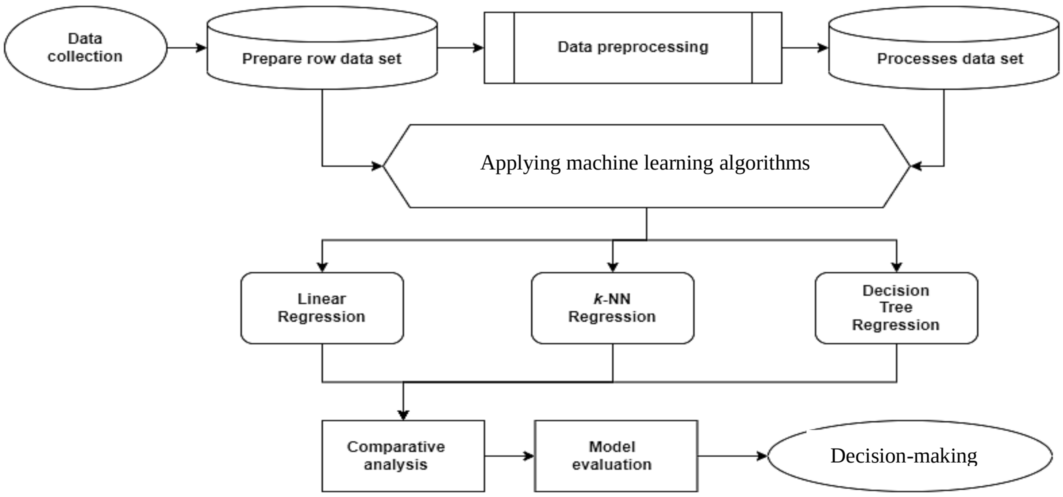

Figure 2. The methodology applied for predicting the change in study performance due to COVID-19

4.1. Description of the data set

We collected data from several universities in Bangladesh. We collected data through various questions and answers. We collected data from students of some private universities with and without COVID-19 about the situation. We collected data based on the twelve features as shown in Tables 1 and 2. The output feature is the SGPA. Thus, the output feature *y* was as follows if the input features were $x_1, x_2, x_3, \ldots, x_{11}$:

$$y = (x1, x2, x3, \ldots, x11) \tag{6}$$

We collected data from 1019 students in total from some private universities in Bangladesh. All of the data we gathered had some numeric data, missing data, noisy values, and some text data, which needed to be dealt with.

4.2. Data preprocessing

We initially checked the data set for missing values before performing level encoding to transform the text data to numerical. The imputer and median were used to fill in the missing data. The data set's noisy values were then evaluated and sorted using the box plot. After that, the normalized data set was processed by applying algorithms to it. This procedure is demonstrated in Figure 3.

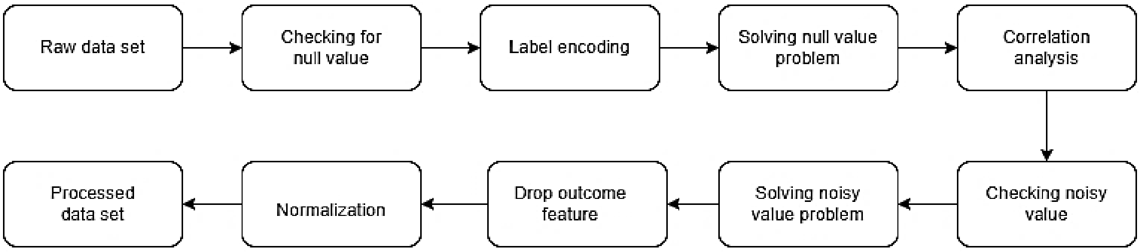

Figure 3. Preprocessing steps for the data that was collected

4.3. Description of the clasification models used

A simple supervised machine learning technique is known as k-NN. The k-NN algorithm can be used to resolve regression and classification issues. The k-NN algorithm checks if nearby neighborhoods include comparable objects [23]. The following (7) is used to calculate the Minkowski distance (D_M) between the query points and other points.

$$D_M=\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \quad (7)$$

Regression and classification are both possible using linear functions. As demonstrated, a linear classifier is produced by passing a linear function's output through a threshold function [23]. The linear regression formula provided is given in (8).

$$Y=a+bX \quad (8)$$

A decision tree represents a function that "decides" a single output value from a vector of input attribute values. Discrete or continuous values can be used for input and output [23]. The decision tree is a classical machine learning algorithm used to model the connection between dependent variables and several independent variables. For now, we will focus on situations where the inputs have discrete values and the result has exactly two potential values. This is a Boolean classification, where each input example will be classified as true or false [23]. Figure 4 provides the decision tree algorithm's working principle.

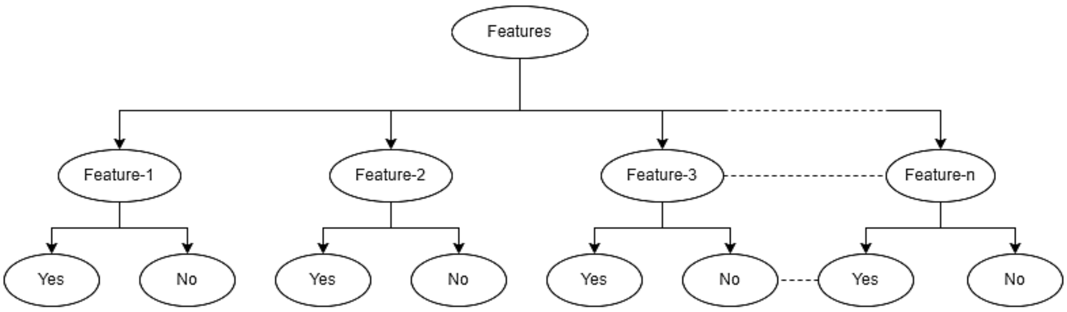

Figure 4. The working principle of the decision tree classification model

5. EXPRERIMENTAL EVALUATION

We divided the entire sample data set, as per the holdout approach [24], into training and testing parts. We apply to determine how much of the data is set aside for training and how much for validation. These data are collected from students studying in different semesters. The data of 1019 students are collected to create a data set. According to the holdout approach, we got the following ratios: train 60% and test 40%; train 70% and test 30%; train 80% and test 20%; train 90% and test 10%. We examine all ratios and find that, of these ratios, training at 80% and testing at 20% gave better performance. We divided the total amount of data we obtained in two, using 80% for training and 20% for testing. In previous sections, detailed discussions of data collecting and pre-processing methods have been provided. Classical machine learning classification models' performance is influenced by the values of several parameters. Table 3 shows the values of some of the three algorithm parameters used in our research.

Table 3. Specifications of the algorithms used in detail

Classification model	Specification
k-NN	Number of neighbors = 1
	Weight function used for prediction, $w(x) = c$, where c is a constant
	Distance metric: Minkowski distance = $\left(\sum_{i=1}^{k}(\lvert x_i - y_i \rvert)^q\right)^{\frac{1}{q}}$
Decision tree	Distribution measure: *Gini index*
	Maximum depth = 0
	Minimum samples split = 2
Linear regression	Fit intercept = 1
	The number of input features = 0
	Copy state of the input features = *True*

Table 4 describes the performance of each algorithm. The algorithms' accuracy, MAE, R^2, MSE, and RMSE performance are evaluated. We would choose the most effective algorithm for our issue area based on the way the approaches worked. We found that the maximum accuracy of linear regression is 99.84%. Accordingly, the best results in terms of MAE, R^2, MSE, and RMSE were likewise successfully produced through linear regression. Good predictions may be obtained through linear regression. The characteristics of linear regression models are well-known and simple to train. Good results are shown by its performance on our dataset. Linear regression works well for prediction models that have a numerical objective.

Table 4. Performance evaluation of the classification models used

Classification Model	Accuracy (%)	MAE (%)	R^2 (%)	MSE (%)	RMSE (%)
k-NN	95.35	0.07	0.21	0.06	0.08
Decision tree	98.78	0.02	0.01	0.02	0.02
Linear regression	99.84	0.01	0.07	0.01	0.01

Based on the method's performance, we will choose the algorithm that is most suitable for our problem area. In comparison to all regressions, it has been found that linear regression performed the best. In conclusion, the linear regression algorithm was used to identify the model that performed better overall.

6. COMPARATIVE ANALYSIS OF RESULTS

We need to compare our work with several current and pertinent research efforts to evaluate the quality of our proposed research performance prediction system based on COVID-19. We should keep in mind that our efforts to evaluate performance will be significantly influenced by the assumptions researchers make when collecting data and presenting the results of their research after processing these samples. We attempted to compare our work to that of others using several criteria, including sample size, feature set size, classification model, and accuracy. Table 5 displays a comparison between our work and other works.

Prakash [4] worked on evaluating COVID-19 datasets and used a random forest classifier to achieve an accuracy of 96.66%. Rai *et al*. [5] used the random forest with 813 students' data and 22 features to analyze student academic performance and achieved a maximum accuracy of 94%. Acharya and Sinha [10] worked on students' performance and achieved 79.0% accuracy using the C4.5 decision tree. Hossain *et al*. [25] worked on online education during COVID-19 with 1,468 students' data and 14 features. Hossain *et al*. [26] found that during COVID-19, researchers in Bangladesh focused on domestic violence and used a random forest to reach an accuracy of 77%. Tarik *et al*. [2] performed a prediction on student performance during COVID-19 with fifteen features after collecting data from 72010 students.

Table 5. Comparison results between our work and other work

Method/work done	Problem dealt with	Sample size	Size of the feature set	Classification model	Accuracy (%)
This work	Change in study performance due to COVID-19	1019	12	Linear regression	99.84
Rai *et al*. [5]	Student academic performance	813	22	Random forest	94
Hossain *et al*. [25]	Prediction of the impact of online education during COVID-19	1468	14	Random forest	99.8
Prakash *et al*. [4]	Evaluation of COVID-19 datasets	Not mentioned	7	Random forest	96.66
Acharya and Sinha [10]	Student's performance	1407	15	Decision tree	79.0
Hossain *et al*. [26]	Domestic violence in Bangladesh during the COVID-19	511	15	Random forest	77

We can see in Table 5 that in this study, we investigated the impact of COVID-19 on study performance, utilizing a sample size of 1019. We analyzed a feature set comprising 12 elements, employing linear regression as our classification model. Our findings revealed a remarkable accuracy of 99.84%, shedding light on the substantial effects of the pandemic on academic outcomes. Concerning such an observation, our acquired accuracy of 99.84% seems to be rather good. The features used in our suggested method provide extremely strong discriminating information to predict the change in study performance due to COVID-19, which is computationally easy to calculate and enables us to obtain very high accuracy. Most of the other works are not particularly similar to our work. Therefore, the merits of our technique should not be openly compared to other works.

7. CONCLUSION AND FUTURE WORK

In this paper, we provide extensive exploratory work to predict changes in research performance due to COVID-19 using various types of machine-learning approaches. Firstly, with the student's input, we created the feature set that serves as the foundation for this predictive work. Preprocessing has been done after the data have been properly evaluated. Three different classifiers have been used to predict the change in study performance caused by COVID-19. Four obvious performance metrics have been used to evaluate the benefits of those classifiers. The relative qualities of the results were then assessed by comparing them to those of related research. Our proposed method has a very good and promising accuracy of 99.84% using linear regression. Educational institutions, students, and their guardians can use this application to get alert and adopt precautionary measures due to the hindrance of any pandemic like COVID-19. A possible future challenge is to work with a much greater dataset of students to cover a broader spectrum of student learning performance required in Bangladesh.

REFERENCES

[1] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 40, no. 6, pp. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.

[2] A. Tarik, H. Aissa, and F. Yousef, "Artificial intelligence and machine learning to predict student performance during the COVID-19," *Procedia Computer Science*, vol. 184, pp. 835–840, 2021, doi: 10.1016/j.procs.2021.03.104.

[3] A. U. Mandayam, R. A.C, S. Siddesha, and S. K. Niranjan, "Prediction of covid-19 pandemic based on regression," in *2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2020, pp. 1–5, doi: 10.1109/ICRCICN50933.2020.9296175.

[4] K. B. Prakash, "Analysis, prediction and evaluation of COVID-19 datasets using machine learning algorithms," *International Journal of Emerging Trends in Engineering Research (IJETER)*, vol. 8, no. 5, pp. 2199–2204, May. 2020, doi: 10.30534/ijeter/2020/117852020.

[5] S. Rai, K. A. Shastry, S. Pratap, S. Kishore, P. Mishra, and H. A. Sanjay, "Machine learning approach for student academic performance prediction," in *Evolution in Computational Intelligence*, vol. 1176, Singapore: Cham, 2021, pp. 611–618, doi: 10.1007/978-981-15-5788-0_58.

[6] M. T. Hossain, N. A. Nishu, A. Sultana, N. Jahan, S. S. Snigdha, and R. Bashitialshaaer, "Challenges of online education/distance learning for university students during the covid-19 pandemic: an observation from Bangladesh," *Research Square*, vol. 1, pp. 1–23, 2021, doi: 10.21203/rs.3.rs-453518/v1.

[7] T. Sultana and R. H. Khan, "Investigating university students' satisfaction on online class: bangladesh perspective," *Bangladesh Education Journal*, vol. 18, no. 2, pp. 23–32, 2019.

[8] M. Tabassum, S. E. Mannan, M. I. Parvej, and F. Ahmed, "Online education during covid-19 in bangladesh: university teachers' perspective," *Aquademia*, vol. 5, no. 1, 2021, doi: 10.21601/aquademia/9611.

[9] Y. Rybarczyk and R. Zalakeviciute, "Assessing the COVID-19 impact on air quality: a machine learning approach," *Geophysical Research Letters*, vol. 48, no. 4, 2021, doi: 10.1029/2020GL091202.

[10] A. Acharya and D. Sinha, "Early prediction of students performance using machine learning techniques," *International Journal of Computer Applications*, vol. 107, no. 1, pp. 37–43, 2014, doi: 10.5120/18717-9939.

[11] J. -P. Vandamme, N. Meskens, and J. -F. Superby, "Predicting academic performance by data mining methods," *Education Economics*, vol. 15, no. 4, pp. 405–419, 2007, doi: 10.1080/09645290701409939.

[12] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in *15th European Concurrent Engineering Conference 2008, ECEC 2008 - 5th Future Business Technology Conference, FUBUTEC 2008*, EUROSIS-ETI, 2008, pp. 5–12.

[13] M. J. Mia, A. A. Biswas, A. Sattar, and M. T. Habib, "Registration status prediction of students using machine learning in the context of private university of bangladesh," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 2594–2600, 2019, doi: 10.35940/ijitee.A5292.119119.

[14] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, 2021, doi: 10.1016/j.crbeha.2021.100053.

[15] M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, and M. T. Habib, "Prediction of addiction to drugs and alcohol using machine learning: a case study on bangladeshi population," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, pp. 4471–4480, 2021, doi: 10.11591/ijece.v11i5.pp4471-4480.

[16] I. F. Rimi, M. A. I. Arif, S. Akter, M. R. Rahman, A. H. M. S. Islam, and M. T. Habib, "Machine learning techniques for dental disease prediction," *Iran Journal of Computer Science (IJCS)*, vol. 5, no. 3, pp. 187–195, Mar. 2022, doi: 10.1007/s42044-022-00101-0.

[17] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[18] J. Xu, K. H. Moon, and M. V. D. Schaar, "A machine learning approach for tracking and predicting student performance in degree programs," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017, doi: 10.1109/JSTSP.2017.2692560.

[19] F. Abdullah, M. A. Rahman, M. Shidujaman, M. Hasan, and M. T. Habib, "Machine learning modeling for reconditioned car selling price prediction," in *Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023)*, SPIE, 2023, doi: 10.1117/12.2689745.

[20] S. Barhmi and O. El Fatni, "Hourly wind speed forecasting based on support vector machine and artificial neural networks," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 3, pp. 286–291, Sep. 2019, doi: 10.11591/ijai.v8.i3.pp286-291.

[21] M. Abdullah, M. Al-Ayyoub, F. Shatnawi, S. Rawashdeh, and R. Abbott, "Predicting students' academic performance using e-learning logs," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 2, pp. 831–839, 2023, doi: 10.11591/ijai.v12.i2.pp831-839.

[22] M. Grari, I. Idrissi, M. Boukabous, O. Moussaoui, M. Azizi, and M. Moussaoui, "Early wildfire detection using machine learning model deployed in the fog/edge layers of iot," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 2, pp. 1062–1073, 2022, doi: 10.11591/ijeecs.v27.i2.pp1062-1073.

[23] S. Russell and P. Norvig, *Artificial intelligence a modern approach*. Upper Saddle River, New Jersey: Pearson Education Limited, 2010.

[24] P. -N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. India: Pearson, 2006.

[25] S. M. Hossain, M. M. Rahman, A. Barros, and M. Whaiduzzaman, "Impact prediction of online education during COVID-19 using machine learning: a case study," in *Intelligent Sustainable Systems*, Springer, vol. 579, Jan. 2023, pp. 567–582, doi: 10.1007/978-981-19-7663-6_54.

[26] M. Hossain *et al.*, "Prediction on domestic violence in bangladesh during the covid-19 outbreak using machine learning methods," *Applied System Innovation*, vol. 4, no. 4, 2021, doi: 10.3390/asi4040077.

BIOGRAPHIES OF AUTHORS

Md. Ataur Rahman completed his graduation from Daffodil International University in computer science and engineering. He enjoys doing research a lot. His papers have previously been presented at global conferences in China and at IEEE. His other papers are being reviewed by a reputable international journal for publication. He has published some of his papers in international conferences and journals. His areas of interest in the study are deep neural networks and machine learning. He has a propensity for conducting research in the areas of IoT and blockchain. He can be contacted at email: ataurrahman862828@gmail.com.

Md. Sadekur Rahman is currently an Assistant Professor in the Department of Computer Science at Daffodil International University. He holds a B.Sc. and an M.Sc. degree in applied mathematics & informatics. With a strong focus on research, he has made significant contributions to the academic community through his publications in various conferences and journals. His research interests center around machine learning, deep learning, and natural language processing. Furthermore, he commitment to advancing knowledge and expertise in his field is evident as he is currently pursuing a Ph.D., further enhancing his research capabilities and contributing to the advancement of the aforementioned areas of study. He can be contacted at email: sadekur.cse@daffodilvarsity.edu.bd.

Mohammad Monirul Islam earned an M.Sc. in computing from London Metropolitan University in the United Kingdom and a B.Sc. in computer science from the University of Pune in India. He is currently pursuing a Ph.D. degree at Jahangirnagar University's Department of Computer Science and Engineering. He is an Assistant Professor at Daffodil International University in Bangladesh, Department of Computer Science and Engineering. He has made major contributions to his profession, with multiple publications in prestigious international journals and conference proceedings. His research interests include machine learning & deep learning, data mining, embedded systems & robotics, and internet of things (IoT) & intelligent systems. He can be contacted at email: monirul.shemul@gmail.com.

Mahady Hasan Ph.D. completed his bachelor degree in computer science from Independent University, Bangladesh in 2001. He received his master's degree in information technology from the University of Sydney in 2004 and his Ph.D. in spatial database from the Department of Computer Science and Engineering at University of New South Wales (UNSW) in 2010. Prior to joining IUB as Assistant Professor, he served UNSW, Australian Pacific College and James Cook University, Australia. He has been working at IUB since 2004. Currently, he is the head of the Department of Computer Science and Engineering and working in the department as Associate Professor. Conducting research in the area of software engineering, big data, data analytics, and engineering education are his key interest. He has published several book chapters, journals and conference papers in many renowned venues. He can be contacted at email: mahady@iub.edu.bd.

Md. Tarek Habib achieved his Ph.D. degree at the Department of Computer Science and Engineering at Jahangirnagar University in 2022. He obtained his B.Sc. degree in computer science from BRAC University in 2006. Then he got an M.S. degree in computer science and engineering from North South University in 2009. He is an Assistant Professor at the Department of Computer Science and Engineering at Independent University, Bangladesh. Earlier, he was an Associate Professor at the Department of Computer Science and Engineering at Daffodil International University. He is much fond of research. He has published many articles in international journals and conference proceedings. His research interest is in artificial intelligence, especially machine learning, artificial neural networks, computer vision, and natural language processing. Recently, he has started working on IoT too. He can be contacted at email: md.tarekhabib@yahoo.com.