International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (Oct.-2021)

http://dx.doi.org/10.12785/ijcds/100187

Machine Learning Approaches to Digital Learning Performance Analysis

Maksud Ahamad[1*], Nesar Ahmad[2]

[1,2] Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Received 15 Mar. 2020, Revised 06 May. 2021, Accepted 13 Jun. 2021, Published 28 Oct. 2021

Abstract: Academic learning performance prediction is one of the concerns for the stakeholders of the educational system, namely administrators, teachers, students, parents, and others. As poor performance in learning may lead to the dropout of a student, so it is vital to predicting the performance to identify the student at risk. By identifying the student at risk, corrective action can be taken in advance for the improvement of the performance. The purpose of this study is to identify the students who may have problems in the coming course sessions. These problems can lead to poor performance. In this study, we have performed comparative analysis for different machine learning algorithms named; Artificial Neural Network (ANN), Naïve Bayes (NB), Decision Tree (DT), Logistic regression (LR), and Support Vector Machine (SVM), on extracted features. The extracted features are average mouse clicks, total activities, average time, average idle time, average keystrokes, and total related activities in an exercise. The results exhibit that SVM is better to predict the performance as equated to other machine learning techniques, by the accuracy of 94.82 %. These findings can suggest measures to take action like additional help required in advance to a particular learner for the success at a higher level of Bloom's Taxonomy.

Keywords: E-learning, Machine Learning, Classification, Support Vector Machine, Deep Learning

1. Introduction

The overall objective of digital learning is to provide a platform for a wider group of people to a higher level of learning outcomes [1]. The term e-learning or digital learning is trending in the current scenario. E-learning is an abbreviation of electronic learning that can be defined as the delivery of learning content (text materials, audio, video, assessment, and interaction methods) through any electronic medium. This can help to improve the level of learning as compared to traditional face-to-face learning. Over the past decade, various variables explored to achieve the goal of performance improvement. These variables are the parameters that may affect student satisfaction and variables that affect the prediction of digital learning outcomes [2].

MOOCs (Massive Open Online Courses) have turn into a substitute educational scenario that permits users the same quality of learning without considering the location of the learners or users [3]. Edx, Coursera, Udacity, HarvardX, Udemy, and Khan Academy are various examples of MOOCs. Because of geographical, financial, and educational obstacles, the use of MOOCs is increasing day by day at an individual as well as institutional level.

In the case of digital learning, where the learners are physically not available, it is essential to track the learners' progress and provide a way of interaction. As most of the digital learning is symmetric for all learners, special care is required to ensure progress for everyone. Machine learning can be a useful technique that can find the hidden patterns from the behavior of users for the performance prediction of the learners[3]. Furthermore, a significant feature of machine learning can explore complex non-linear patterns in the users' behavior[4]. These machine learning methods can identify the critical student who could face some difficulties in succeeding course learning [5].

Educational Data Mining (EDM) and learning analytics have gained attention among the user of TEL (Technology Enhanced Learning) during previous years [6] [25]. EDM explores the educational data and tries to find the learning behavior to help the learners and the

E-mail:maksud.ahmad12@gmail.com, nesar.ahmad@gmail.com

teachers or instructors [7] [8]. In this work, we have used machine learning algorithms to predict the students' or learners' performance in advance to take corrective actions. For this, the features which are extracted from the dataset used in this study are the inputs to train and test the machine learning algorithms. These features are extracted from the dataset that is discussed below in this paper. The model that is created with the previous session data can predict the learner's performance for the selected features of the next course session. This early prediction may be beneficial for learners, teachers, and other stakeholders.

The rest of the paper is organized into another four sections. Section 2 has discussed the related work. Section 3 addresses the methods used in this work. Results and discussion are comprised in section 4. Finally, the conclusion is included in section 5.

2. Related Works

Foreseeing low-commitment learners is fundamental in e-learning frameworks since it enables educators to comprehend the conduct of students for various course exercises. Here the machine learning techniques are applied and evaluated with cross-validation methods. Analysis indicated that J48, DT (Decision Tree), JRIP, and Gradient Boosting Trees (GBT), are the best methods to foreseeing low-commitment learners or students during an Open University evaluation [9].

Vahdat et al. (2015) utilized the complexity matrix (CM) and process mining (PM) strategies to examine the connection among student's learning stages and grades with Digital Electronics Education and Design Suite (DEEDS), an electronic learning framework. They inferred that the average learners' grades decidedly correspond with the complexity matrix, and the difficulty is contrarily related to the complexity matrix[10]. For the learners' or students' performance issues, a method is proposed, namely GritNet, which expands upon the bidirectional long short term memory (BLSTM) [11].

Abu Saa (2016) experimented with locating the best classifier to predict learners' performance in higher education, utilizing social and individual information highlights [12]. In order to forecast student's performance by logging information during student interactions, several probability models (i.e., Bayesian tracking) are used. However, the hidden behaviors of students are not predicted by these models [13].

The study compared the EDM (educational data mining) methods (Naive Bayes, Neural Network, and Support Vector Machine) [14]. The findings achieved here, demonstrating that the analyzed EDM procedures are adequately capable of early recognize students' failure. So these methods are valuable to furnish instructors or educators with essential data to support their choices.
Various classifier-based systems are exhibited in [15], which bring together three classifiers, namely, AODE, IBK, and J48, using the democratic philosophy and presented a solitary composite model. In this study, they proposed a technique for anticipating the last grades of students by a Recurrent Neural Network (RNN) from the log information put away in the instructive frameworks. The exactness of forecast by the RNN is above 90% utilizing the log information until the sixth week [16].

3. Methodology

In this section, we have described the dataset, performance evaluator or measures, features extraction, features selection, and the machine learning algorithms used in this study.

A. Dataset

The data is collected from the undergraduate students of the University of Genoa [10]. There are a total of 100 students participated in the study. The data was recorded using a simulator termed Digital Electronics Education and Design Suite (DEEDS). DEEDS is used for e-learning in the digital electronics course. The system provides content for learning through particular browsers, and the students are requested to take care of different issues with various degrees of challenges regarding the course content. This data comprises the participants' logs of activities held over the completion of the digital electronics course. There were a total of six lab sessions throughout the course. The data from every session is organized as, Id of the session, Id of student, Id of the exercise, Labeled activity, Activity start time, Activity end time, Idle time, Mouse wheel amount, Mouse wheel clicks, Mouse clicks left, Mouse clicks right, Mouse movements, and Keystrokes numbers. There are various activities held during the completion of the exercise, like reading or viewing the material concerning or not concerning the exercise, working on DEEDS, working on the text editor, doing nothing, and students using the learning management system.

B. Performance Evaluators

The two types of assessments identified and compared with the outcomes of the performance evaluation:
Visualization: The result of true values and false values can be represented for classifiers using the receiver operative curve (ROC) and reject curves.
Statistical Analysis**:** The outcomes from different classification methods are evaluated using confusion matrix (accuracy), precision, recall, and F-measure. The mathematical formulae for evaluators used in this study are as follows,

Confusion Matrix: This displays the errors through all classes. The accuracy of the classifier is determined as follows [5]:

Accuracy= (TP (True Positive) + TN (True Negative)) / (TP+TN+FP+FN)*100%

TABLE I. MEANING OF TP, TN, FN, AND FP.

		Actual	
		Positive	***Negative***
Prediction	***Positive***	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Precision: It is a ratio of TP (true positive) and the misclassified values as positive (e.g., FP).

Precision= TP/ (TP+FP)

Recall: Recall is a proportion of correctly categorized items, i.e., TP, and the values that are categorized wrongly, i.e., false negative.

Recall/Sensitivity= TP / (TP +FN)*100%

F-Measure: It is another standard output measure that combines recall and precision into a single measure [17]. The formula used to calculate F-measure is as follows:

F-Measure = 2*(precision*recall)/ (precision + recall)

ROC curve: It includes the error types I and II. The values of "false positives" and "false negatives" are represented.

$$\text{Type I Error} = \frac{FN}{TP+FN} \quad \text{and}$$

$$\text{Type II Error} = \frac{FP}{TN+FP}$$

C. Feature Extraction

The process of transformation of original features to more valuable features is called feature extraction. Brian Ripley defines feature extraction as the construction of the linear combination of continuous features that can discriminate classes[6].

In this study, we have used Spyder, an open-source, cross-platform IDE (integrated development environment) for implementing this work. Here, we have extracted the thirty-six features for each session from the dataset that is described above. As there are six exercises in a session, so we got thirty-six features for every student. We can see the extracted features in table II. All the features are described as follows,

The average time of the exercise is the difference between the beginning and the end to complete an exercise. Users that spend less time on the current session are not well oriented [24]. Overall activities are the total no of activities during the completion of an exercise. A participant who completes a broader range of activities while answering the given questions may affect the student's performance. The third feature is average mouse clicks during the completion of an exercise. The number of mouse clicks may lead to a pattern that relates to the performance of students. The fourth feature is the average of spent idle time throughout an exercise. The students who spent more time in do nothing may have some problems. The fifth feature is described as keystroke average during the completion of an exercise. The average keystrokes relate to the engagement of students. The student achieved better grades who had more keystrokes [24]. The last feature is overall related activities over an exercise completion. If a student using fewer activities related to the exercise may have problems in the future.

TABLE II. LIST OF RAW INFORMATION AND EXTRACTED FEATURES

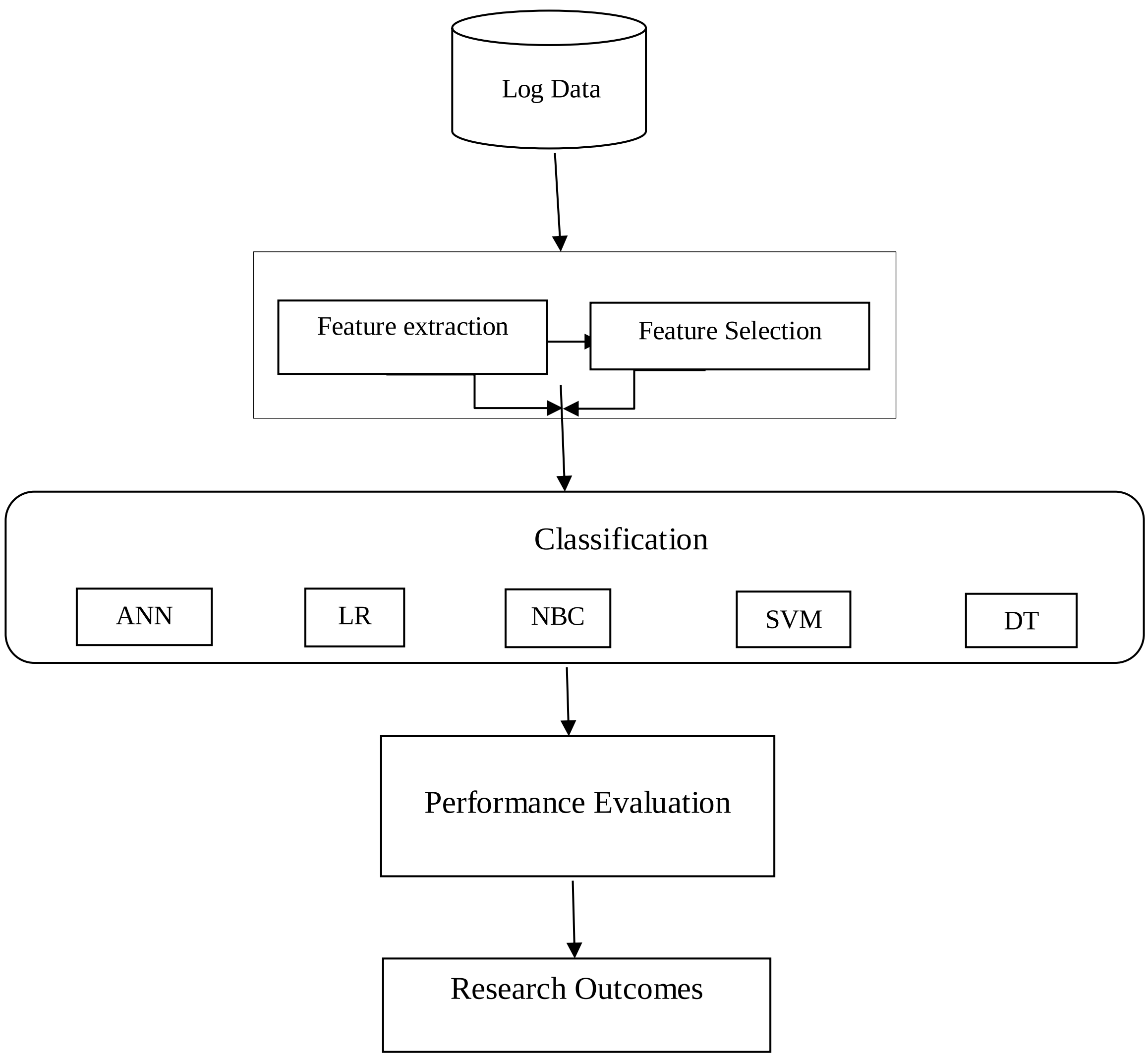
Raw Information	**Extracted Features**
Id of session Id of student Id of the exercise Labeled activity Activity start time Activity end time Idle time Mouse wheel amount Mouse wheel clicks Mouse clicks left Mouse clicks right Mouse movements Keystrokes numbers	➢ Average time in the exercise ➢ Overall activities in the exercise ➢ Average mouse clicks in the exercise ➢ Average of the idle time in the exercise ➢ Keystrokes' average in the exercise ➢ Overall related activities in the exercise

D. Feature Selection

A feature or attribute selection is the method of selecting the best attributes or features from the given set of features, as some features may be statistically uncorrelated with the class, target, or label[8]. Some features may be redundant or misleading in the case of high-dimensional data. This redundancy or irrelevancy of features may affect accuracy and efficiency [9].

We have selected the attribute or feature which has the highest ranking score. In this study, we have selected three evaluators from the Waikato Environment for Knowledge Analysis (WEKA) tools. These are the Gain Ratio Attribute Evaluator, Chi-squared Attribute

Evaluator, and Info Gain Attribute Evaluator. The Gain Ratio tests the attribute with the class using the gain ratio value[10]. The Chi-squared Attribute Evaluator tests the attribute with the class using the chi-squared statistic. The Info gain attribute evaluates the features based on information gain [21].



E. Algorithms Used

Here, we are discussing the details of the various machine learning algorithms used for the training and testing data.

Artificial Neural Network (ANN): Usually used in the neural network for classification is Multilayer Perceptron (MLP). During simulations with the data set, the MLP architecture includes a network of three layers, one input, hidden, and output layer. In this model, the parameters used are, learning rates= 0.3, threshold for validation= 20, momentum= 0.2, Number of Epochs = 500, random seed=0 [22].
Logistic regression (LR): The application of logistic regression applies multiple regression analysis techniques to the cases in which the output variable is categorized. The relation between the attribute and classifier is a logistic regression function rather than a linear relation. In logistic regression, the dependent variable (class) is binary, indicating dependent attributes will produce 1 with a probability of success pi or 0 with 1-pi with the probability of failure [23].

Naïve Bayes (NB): NB (Naive Bayes) provides an easy solution to probabilistic thinking. It relies on two hypotheses that the prediction attributes are an independent class label, and the prediction is regulated by suppressed attributes [15].

Support Vector Machines (SVM): The SVM is also referred to as the highest margin classifier, which optimizes the distance between the support vectors and hyperplane. Such vectors are used as learning vectors nearby to the hyperplanes from each group or class. Using kernel function, SVM can classify linear and non-linear data [16]. Here we are using the SMO (Sequential minimum optimization) for SVM (Support vector machine).

Decision Tree (DT): The DT is a set of internal nodes and leaves. The internal node that can have two or more child nodes represents attributes or features of a dataset. Moreover, the values indicated by the branches and the class represented by the leaf node [26]. A DT is essential when a study tries to identify which features in a student performance prediction model are significant. From the data set, the tree developed, indicating what information at the child node is best divided [27].

4. Results and Discussion

To assess the efficacy of this study, we have performed some evaluations using the performance evaluators mentioned above. In this work, our target is to predict the problems that learners will face in a succeeding course session. By identifying the problems in advance, we can predict the performance of a particular student. The dataset used here is collected from undergraduate engineering students of the University of Genoa [10]. Here, the Spyder IDE is used to extract the feature from the raw data collected from the DEEDS simulator.

For this work, initially, we have extracted the no of features from the raw data collected from the undergraduate students using the e-learning platform. The features which we have extracted are the average time in the exercise, overall activities in the exercise, average mouse clicks in the exercise, average of idle time in the exercise, keystrokes' average in the exercise, and overall related activities in the exercise. These extracted features are used as input variables for training machine learning algorithms. Moreover, the grades which students got in the sessions were used as output variables or class variables. The grades of students are divided into two groups, which are used as the output variable. The students who get a grade of less than two will have the problems in coming sessions and the rest of the grades (grades$\geq$2) students. After feature extraction, we performed some preprocessing steps because it is essential before using the machine learning methods. In the feature extraction phase, we are normalizing the extracted features to get the same scaled values.

To evaluate the effectiveness of the various extracted features from the dataset, here we have performed various experiments with the help of classification methods (ANN, LR, NBC, SVM, and DT). For the first experiment, datasets collected from the various sessions were divided into the 80:20 ratio to perform the training and testing. We are dividing the datasets into this ratio because it gives better results [28]. The training data includes a total of three sixty-one records from the first four sessions, and the total eighty-five records from the last session are used for the testing purpose. Subsequently, dividing the data into training and testing sets, we have used ANN, LR, NBC, SVM, and DT, to train the model with the training data. Now the new data is used to test the model.

In the first experiment, we have experimented five times with each algorithm as there are five sessions and found the average accuracy, RMSE, Precision, Recall, and F1-score. As we can see from table III and Fig. 2, we are getting better results in terms of accuracy, RMSE, Recall, Precision, and F1-score as compared to M. Hussain et al. [24]. Here we have used an extra feature, that is, Average

Mouse clicks in the exercise as compared to the M. Hussain et al. [24]. We can see from table III that SVM got the highest accuracy as compared to the ANN, LR, NBC, and DT algorithms.

The ROC (Receiver operative characteristic) curve is also plotted for the values we found in the experiments. Here we have plotted the false positive rate against the true positive rate, which we found with different algorithms. From Fig. 3, we can draw that the SVM got better ROC values as compared to the other algorithms that mean better results.

TABLE III. COMPARISON OF OUTCOMES OF ANN (ARTIFICIAL NEURAL NETWORK), LR (LOGISTIC REGRESSION), NBC (NAIVE BAYES CLASSIFIERS), DECISION TREE (DT), AND SVM (SUPPORT VECTOR MACHINE), WITH ALL THE FEATURES

Classifier	Average Accuracy		Average RMSE		Average Precision		Average Recall		Average F1-Score	
	M. Hussain et al.[24]	Proposed Work	M. Hussain et al.[24]	Proposed Work	M. Hussain et al.[24]	Proposed Work	M. Hussain et al.[24]	Proposed Work	M. Hussain et al.[24]	Proposed Work
ANN	75.00	81.26	0.48	0.31	0.80	0.91	0.91	0.81	0.85	0.83
LR	73.00	81.60	0.50	0.32	0.79	0.89	0.90	0.81	0.84	0.83
NBC	75.00	72.43	0.49	0.41	0.82	0.91	0.90	0.72	0.85	0.72
SVM	75.00	89.11	0.48	0.24	0.80	0.88	0.91	0.88	0.85	0.88
DT	69.00	88.37	0.54	0.28	0.79	0.89	0.83	0.88	0.81	0.88

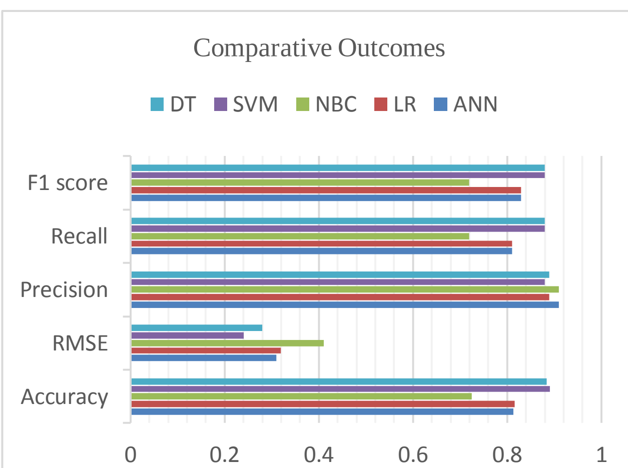


Figure 2: Accuracy, Root Mean Squared Error (RMSE), Precision, Recall and F1 score for ANN (artificial neural network), LR (logistic regression), NBC (Naive Bayes classifiers), decision tree (DT), and SVM (support vector machine) algorithms with all the features

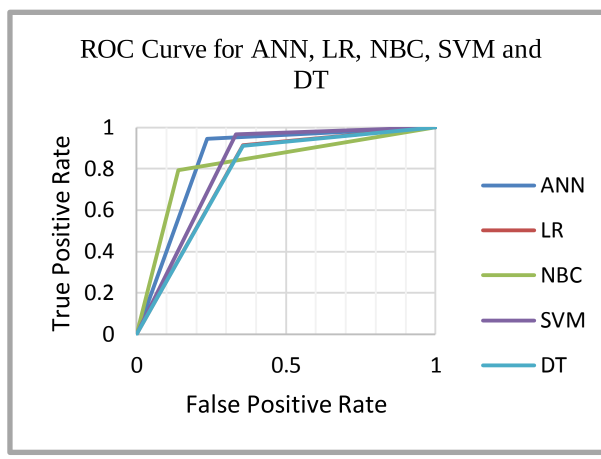


Figure 3: Receiver operator characteristic (ROC) curves for the ANN, LR, NBC, and SVM algorithms with all the features

In the second evaluation, we have used feature selection methods that reduce the complexity by reducing the dimension of data. Here we are selecting the features from the total numbers of features used in the first experiment of this study. We have selected the features using the majority voting techniques with all the evaluators. Here we are using the Gain Ratio Attribute Evaluator, Chi-squared Attribute Evaluators, and Info Gain Attribute Evaluator. Finally, with these three feature selection methods, we have taken the best ten features from thirty-six features. Now with these selected features, the algorithms (ANN, NBC, LR, DT, and SVM) are trained using the data selected for training and tested with the data used for testing. Here we have applied the five-fold cross-validation technique. We found the best results for the decision tree as compared to the other classifiers. We can see the results from table IV and Fig. 4.

TABLE IV. COMPARISON OF OUTCOMES OF ANN (ARTIFICIAL NEURAL NETWORK), LR (LOGISTIC REGRESSION), NBC (NAIVE BAYES CLASSIFIERS), DECISION TREE (DT), AND SVM (SUPPORT VECTOR MACHINE) WITH THE SELECTED FEATURES

Classifier	Accuracy	RMSE	Precision	Recall	F1-Score
ANN	92.76	0.24	0.92	0.92	0.92
LR	92.53	0.23	0.92	0.92	0.92
NBC	87.10	0.32	0.92	0.87	0.88
SVM	92.53	0.27	0.92	0.92	0.90
DT	93.43	0.24	0.93	0.93	0.93

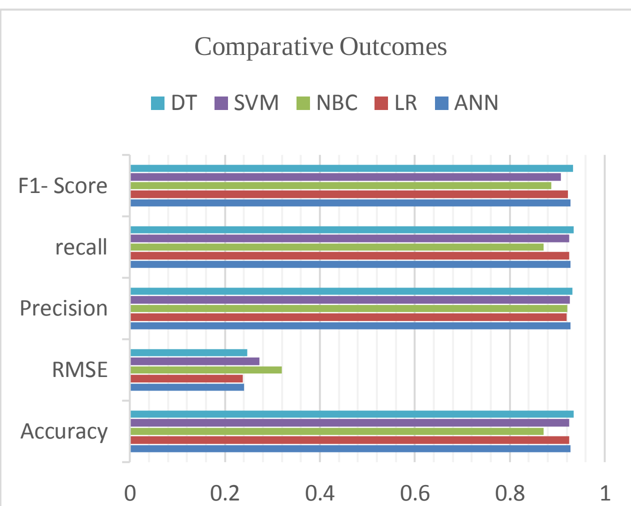


Figure 4: Accuracy, Root Mean Squared Error (RMSE), Precision, Recall and F1 score for ANN (artificial neural network), LR (logistic regression), NBC (Naive Bayes classifiers), decision tree (DT), and SVM (support vector machine) algorithms with the selected features

In the third experiment, we are replacing the overall idle time with the average mouse clicks in the exercise. We are replacing this feature to keep the equal no of features as in M. Hussain et al. [24]. The achieved outcomes are in table V and plotted in Fig. 5. In the last experiment, we have used basic deep learning concepts and found improvement as compared to the previous experiments. Here we got the 94.38 percent accuracy for 200 epochs, which we see from table VI and Fig. 6.

TABLE V. RESULTS AFTER REPLACING A FEATURE (AVERAGE IDLE TIME) WITH THE OTHER FEATURE (AVERAGE MOUSE CLICKS)

Classifier	Accuracy	RMSE	Precision	Recall	F1-Score
ANN	92.24	0.21	0.93	0.92	0.92
LR	77.46	0.34	0.92	0.77	0.81
NBC	82.53	0.34	0.92	0.82	0.83
SVM	94.82	0.13	0.94	0.94	0.94
DT	92.65	0.18	0.93	0.92	0.93

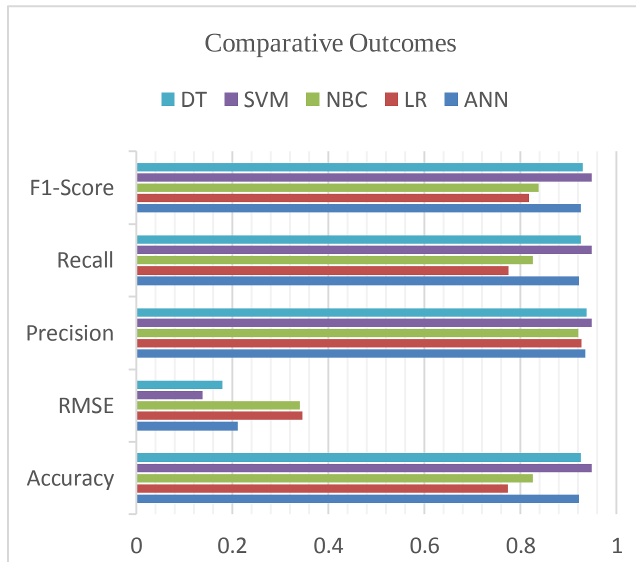


Figure 5: visualization of results after replacing a feature (average idle time) with the other feature (average mouse clicks)

TABLE VI. RESULTS WITH THE DEEP LEARNING TECHNIQUE

Deep Learning (With four hidden layer, RMSprop Optimizer, and 100 nodes in each hidden layers)	Accuracy	Precision	Recall	F1-Score
No of epoch=100	91.01	0.92	0.97	0.96
No of epoch=200	94.38	0.95	0.98	0.96
No of epoch=300	88.76	0.94	0.92	0.93
No of epoch=400	89.88	0.92	0.96	0.94
No of epoch=500	93.25	0.95	0.97	0.96
No of epoch=1000	92.13	0.95	0.96	0.95

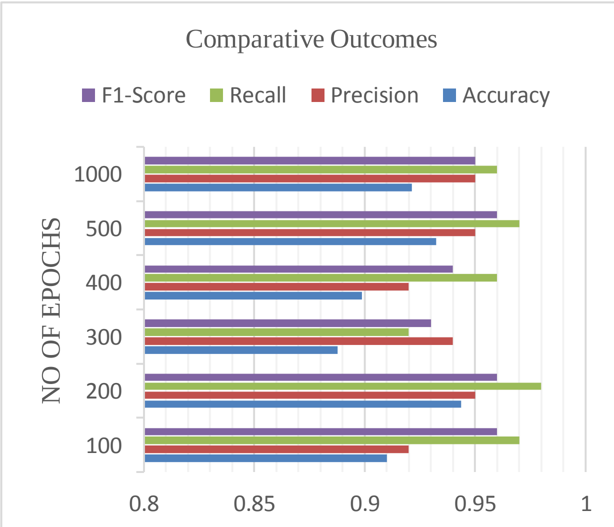


Figure 6: visualization of results with no of epochs

5. CONCLUSION

In this work, we have conducted some experiments to predict the students' problems that may lead to poor performance, in advance to take corrective action.

Initially, we have extracted the features from the DEEDS logged data. After the extraction of features from raw data, we have taken the data from the previous course sessions data to train the classifiers (ANN, NBC, LR, DT, and SVM) and tested the algorithms on the new course session data.

In this study, we have performed the first experiment by dividing the datasets into an 80:20 ratio and took the average of all the results for every classifier. For this experiment, to predict the performance, SVM attained the highest accuracy, which is 89.11%, as compared to the other algorithms. This is illustrated by the ROC curve also. In the next experiment, we have used three attribute selection methods to select the attributes or features from all the extracted features and selected ten features using a democratic way. For this experiment, the Decision Tree achieved the highest accuracy as compared to the other classifiers for the five-fold cross-validation method.

Additionally, we have replaced a new extracted feature (Average mouse clicks) with existed average idle time features and achieved 94.82% accuracy for SVM. In the last experiment, we have used the deep learning concept and achieved better results as compared to the first experiment of this study that includes all the extracted features. We can use these types of findings in digital learning as well as in academics to overcome the problems of failure of the students. In the future, we can use the ensemble algorithms to improve the performance of the proposed work as ensemble methods take the decision by considering more than one view at a time.

ACKNOWLEDGMENT

This work was supported by the Visvesvaraya Ph.D. Scheme for Electronics and IT fellowship of Ministry of Electronics and Information Technology (Meity), Government of India.

REFERENCES

[1] M. H. Lin, H. C. Chen. and K. S. Liu, "A study of the effects of digital learning on learning motivation and learning outcome." Eurasia Journal of Mathematics, Science and Technology Education 13, no. 7, 2017, 3553-3564.

[2] H. M. S. Ahmed, "Hybrid E-Learning acceptance model: Learner perceptions." Decision Sciences Journal of Innovative Education 8, no. 2, 2010, 313-346.

[3] J. Qiu, T. Jie, T. X. Liu, G. Jie, Z. Chenhui, Z. Qian, and Y. Xue. "Modeling and predicting learning behavior in MOOCs." In Proceedings of the ninth ACM international conference on web search and data mining, ACM, 2016, pp. 93-102.

[4] D. Gasevic, C. Rose, G. Siemens, A. Wolff, and Z. Zdenek, "Learning analytics and machine learning." In Proceedings of the Fourth International Conference on learning analytics and knowledge, ACM, 2014, pp. 287-288.

[5] A. Tamhane, I. Shajith, B. Sengupta, M. Duggirala, and J. Appleton. "Predicting student risks through longitudinal analysis." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 1544-1552.

[6] G. Siemens. "Learning analytics: envisioning a research discipline and a domain of practice." In Proceedings of the 2nd international conference on learning analytics and knowledge, pp. 4-8. ACM, 2012.

[7] R. Baker, Shaun, and P. S. Inventado. "Educational data mining and learning analytics." In Learning Analytics, pp. 61-75. Springer, New York, NY, 2014.

[8] J. Fiaidhi. "The next step for learning analytics." IT Professional 16, no. 5 (2014): 4-8.

[9] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi. "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores." Computational intelligence and neuroscience 2018.

[10] M. Vahdat, O. Luca, A. Davide, F. Mathias, and R. Matthias. "A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator." In Design for teaching and learning in a networked world, pp. 352-366. Springer, Cham, 2015.

[11] B. H. Kim, E. Vizitei, and V. Ganapathi, "GritNet: Student performance prediction with deep learning." arXiv preprint arXiv: 1804.07405 2018.

[12] A. A. Saa, "Educational data mining & students' performance prediction." International Journal of Advanced Computer Science and Applications 7, no. 5, 2016, 212-220.

[13] T. Käser, R. Nicole. Hallinen, and I. S. Daniel. Schwartz. "Modeling exploration strategies to predict student performance within a learning environment and beyond." In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 31-40. ACM, 2017.

[14] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in

introductory programming courses." Computers in Human Behavior 73, 2017, 247-256.

[15] M. Pandey, and S. Taruna. "Towards the integration of multiple classifier pertaining to the Student's performance prediction." Perspectives in Science 8, 2016, 364-366.

[16] F. Okubo, Y. Takayoshi, A. Shimada, and O. Hiroaki. "A neural network approach for students' performance prediction." In LAK, pp. 598-599. 2017.

[17] M. Sokolova, and L. Guy, "A systematic analysis of performance measures for classification tasks." Information processing & management 45, no. 4, 2009, 427-437.

[18] Y. Zheng, V. Brian, D. Ebenezer, S. Dwight, M. Maureen, D. Brainard, and J. Gee. "An automated drusen detection system for classifying age-related macular degeneration with color fundus photographs." In 2013 IEEE 10th International Symposium on Biomedical Imaging, pp. 1448-1451. IEEE, 2013.

[19] J. Han, P. Jian, and M. Kamber. Data mining: concepts and techniques. Elsevier, 2011.

[20] S. Khalid, T. Khalil, and N. Shamila. "A survey of feature selection and feature extraction techniques in machine learning." In 2014 Science and Information Conference, pp. 372-378. IEEE, 2014.

[21] I. H. Witten, F. Eibe, A. H. Mark, and J. P. Christopher, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[22] L.I. Kuncheva, Combining pattern classifiers: methods and algorithms. John Wiley & Sons, 2014.

[23] K. Sotiris, P. Christos, and P. Panagiotis. "PREDICTING STUDENTS'PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES." Applied Artificial Intelligence 18, no. 5, 2004, 411-426.

[24] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and Sadaqat Ali. "Using machine learning to predict student difficulties from learning session data." Artificial Intelligence Review, 52, no. 1, 2019, 381-407.

[25] F. Albalooshi, H. AlObaidy, and A. Ghanim, “Mining Students Outcomes: An Empirical Study,” International Journal of Computing and Digital Systems, 8(03), 2019, pp.229-241.

[26] R. R. Kabra, and R. S. Bichkar, "Performance prediction of engineering students using decision trees." International Journal of computer applications 36, no. 11, 2011, 8-12.

[27] R. Asif, M. Agathe, and M. K. Pathan. "Predicting student academic performance at degree level: a case study." International Journal of Intelligent Systems and Applications 7, no. 1, 2014, 49.

[28] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation." (2018).

Maksud Ahamad received the M.Tech degree from the Indian Institute of Information Technology (IIIT) Allahabad, India, in 2011. He is currently working as a research scholar in the department of computer engineering, Zakir Husain College of Engineering & Technology, Aligarh Muslim University, Aligarh, India. His interests are Soft Computing based machine learning algorithms, Artificial Intelligence, Natural Language Processing, and Data Mining.

Nesar Ahmad received B.Sc (Engg) degree in Electronics & Communication Engineering from Bihar College of Engineering, Patna, India (Now NIT, Patna) in 1984. He received M.Sc (Information Engineering) degree, City University of London, U.K., in 1988, and a Ph.D. degree from the Indian Institute of Technology, Delhi, India, in 1993. He is currently a Professor in the Department of Computer Engineering, Zakir Husain College of Engineering & Technology, Aligarh Muslim University, Aligarh, India. His current research interests mainly include Applied Soft Computing, Digital Learning, and Natural Language Processing.