Webology (ISSN: 1735-188X)
Volume 18, Number 5, 2021

An Auto-Encoder Based Feature Selection Method For Classification Of Student Performance

K. LAKSHMI[1] , Dr. M. RAJA KUMAR[2]

[1]Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

[2]Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

ABSTRACT

Educational organizations are unique and play utmost significant role for the development of any country. As Education transforms the lives of individuals, families, communities, societies, countries and ultimately the world! This is why we live comfortable lives today. Data Mining is the most prevalent techniques to evaluate students' performance and is extensively used in educational sector known as educational data mining. It is evolving area of study that emphases on various techniques of data mining like classification, prediction, feature selection. It is employed on learning records or data related to education to predict the students' performance and learning behaviour by extracting the hidden knowledge. In this research work, Deep learning technique-based feature selection method is proposed in the pre-processing stage to enhance the prediction accuracy of the student performance.

KEYWORDS: Student Performance Prediction, Data Mining, Machine Learning, Deep Learning, Educational Data Mining, Feature Selection, Autoencoder

1. INTRODUCTION

Economic growth is positively correlated with the level of education. To support this growth, educational systems need to improve their students' learning process [1]. Educational systems for universities, schools, and training centers have a large volume of complex data (information) that can be analyzed and used to extract meaningful knowledge that could support the decision-making process for enhancing the educational organizations [2] [3]. However, the ability of human experts in analyzing and processing the vast amount of data collected from educational systems and exploiting this data to extract new useful information is limited. Thus, this large amount of data requires a computerized approach to process them. In this work, Educational Data Mining approach (EDM) is used.

EDM is an area of employing different datamining methods (i.e., clustering, classification, and statistical methods) on data collected from various resources in educational systems, such as admissions/registration data, student marks, virtual courses, and e-learning log files [4], to discover the hidden knowledge [5]. The extracted knowledge can help tutors and decision-makers to take the correct decisions for solving educational issues in their organizations in an appropriate manner. Moreover, in educational systems, EDM and learning analytics (LA) are two main components to collect, analyze, and report all educational data to enhance the educational process. In general, EDM can be categorized into two main parts (i) applied research, where the extracted knowledge focuses on enhancing the learning quality, and (ii) pure research, where the extracted knowledge focuses on increasing our understanding of the learning process [6].

Machine learning (ML) has been employed successfully in the educational field. Educational systems such as Learning Management Systems (LMS) can hold a huge amount of data year after year. This data should be analyzed, classified, or clustered correctly using ML methods to understand the hidden knowledge and differentiate between excellent, good, and weak students [7]. Determining students' performance will encourage excellent students to keep their high performance, while weak students may be given great attention by tutors to improve their performance [8]. For example, in LMS, there are many activities such as homework, quizzes, and virtual classes. These activities may need some prediction models using ML, such as prediction of quiz failure, submitting homework after the due date, or students unable to complete a virtual class. Prediction systems using ML methods will give tutors a better understanding of their students' performance in advance [9] [10].

2. RELATED WORKS

Turabieh, Hamza, et al [11] proposed a modified version of Harris Hawks Optimization (HHO) algorithm by controlling the population diversity to overcome the early convergence problem and prevent trapping in a local optimum. The proposed approach is employed as a feature selection algorithm to discover the most valuable features for student performance prediction problem. A dynamic controller that controls the population diversity by observing the performance of HHO using the k-nearest neighbors (kNN) algorithm as a clustering approach. Once all solutions belong to one cluster, an injection process is employed to redistribute the solutions over the search space. A set of machine learning classifiers such as kNN, Layered recurrent neural network (LRNN), Naïve Bayes, and Artificial Neural Network are used to evaluate the overall prediction system.

Turabieh, Hamza [12] applied a hybrid feature selection algorithm with different machine learning classifiers (i.e. nearest neighbors (kNN), Convolutional Neural Network (CNN), Naive Bayes (NB) and decision trees (C4.5)) to predict the student's performance. A feature selection algorithm is used to select the most valuable features. The authors applied a binary genetic algorithm as a wrapper feature selection.

Gajwani, Juhi, and Pinaki Chakraborty [13] proposed a method to classify a student's performance based on a subset of behavioural and academic parameters using feature selection and supervised machine learning algorithms such as logistic regression, decision tree, naïve Bayes classifier and ensemble machine learning algorithms like boosting, bagging, voting and random forest classifier. For selection of the attributes, we plotted various graphs and determined the attributes that were most likely to affect and improve prediction. The authors predicted the academic performance of a student based on certain attributes of an educational dataset.

Sultana, Jabeen, M. Usha, and M. A. H. Farquad [14] discovered the performance of students using some classification techniques and discovering the best one which yields optimal results. Educational Dataset is collected from a Saudi University database. The dataset is pre-processed to filter duplicate records; missing fields are identified and filled with the destined data. Deep Learning techniques like Deep Neural Net and Data Mining techniques like Random Forest, SVM, Decision Tree and Naïve Bayes are employed on the data set using Weka and Rapid Miner tools.

Sokkhey, Phauk, et al [15] used the spot-checking algorithm to compare these methods and find the most effective method. The authors proposed three main classes of education research tools: a statistical analysis method, machine learning algorithms, and a deep learning framework. The data were obtained from many high schools in Cambodia. The authors introduced feature selection techniques to figure out the informative features that affect the future performance of students in mathematics. The proposed ensemble methods of tree-based classifiers provide satisfiying results, and in that, random forest algorithm generates the highest accuracy and the lowest predictive mean squared error, thus showing potential in this prediction and classification problem.

Nagy, Marcell, and Roland Molontay [16] employed and evaluated several machine learning algorithms to identify students at-risk and predict student dropout of university programs based on the data available at the time of enrollment (secondary school performance, personal details). The authors also presented a data-driven decision support platform for education directorate and stakeholders. The models are built on data of 15,825 undergraduate students from Budapest University of Technology and Economics enrolled between 2010 and 2017 and finished their undergraduate studies either by graduation or dropping out. The authors handled the problem of missing data by imputation. After performing feature extraction and feature selection, a wide range of classifiers have been trained including Decision Tree-based algorithms, Naive Bayes, k-NN, Linear Models and Deep Learning with different input settings.

Naicker, Nalindren, Timothy Adeliyi, and Jeanette Wing [17] investigated support vector machines has been used extensively in classification problems; however, the extant of literature shows a gap in the application of linear support vector machines as a predictor of student performance. (e aim of this study was to compare the performance of linear support vector machines with the performance of the state-of-the-art classical machine learning algorithms in order to determine the algorithm that would improve prediction of student performance. In this quantitative study, an experimental research design was used.

Arif, Md, et al [18] proposed the feature selection algorithm and classification model to develop student performance prediction system. The performance of the proposed methods is compared with another feature selection algorithm that is based on classification model. The authors aimed to provide machine learning classifier algorithms with selected attribute to get better accuracy and comparable different feature selection algorithms because of its feature selection algorithm which is used to select the most valuable features. In the feature selection algorithm, Wrapper Subset Eval method with Random Forest classification is better than other feature selection method and other classification algorithm to predict students' performance.

Hidalgo, Ángel Casado, Pablo Moreno Ger, and Luis De La Fuente Valentín [19] explored the potential of Deep Learning and Meta-Learning in this field, which has thus far been explored very little, so that it can serve as a basis for future studies. The authors implemented a predictive model which is able to automatically optimise the architecture and hyperparameters of a deep neural network, taking as a use case an educational dataset that contains information from more than 500 students from an online university master's degree.

3. PROPOSED AUTO-ENCODER (AE) BASED FEATURE SELECTION (FS) METHOD

The proposed AE-FS method uses Mutual Information technique to calculate the probability of two features occurs at the same time and Stacked Autoencoder for extracting the feature vector. For the input representation layer, let the number of features in the dataset be |V | and the dimension of a vector representing features be v << |V |, where v is the number of features that occur in the text more than a threshold frequency t^2. The idea is to represent feature depending on the most informative target. Therefore, the matrix of all feature vectors is noted as $M \in R^{|V| \times v}$. Each row in matrix M corresponds to the representative vector of feature w_i. We associate to the j-th element of f_i's vector, the value of MI (f_i, f_j), where $j \in \{1, .., v\}$. MI is one of the most extensively used feature selection criteria. It is an information measure that indicates which features tend to often co-occur in a context. It measures the statistical correlation between pairs of features, which is formulated by (1),

$$\mathrm{MI}(f_i, f_j) = \log \frac{P(f_1, f_2)}{P(f_1)P(f_2)} \qquad (1)$$

Where f_1 and f_2 are two features, $P(f_1, f_2)$ is the probability that both features co-occur at the same time, and $P(f_1)$ and $P(f_2)$ are the probabilities that an isolated word occurs. The ratio shows a metric of statistical dependence between f_1 and f_2. Therefore, the MI measures determines the degree of co-occurrence between two features, f_1 and f_2. In fact, this measure is very high, that is, when f_1 and f_2 are often found together, they are dependent, i.e., the existence of one depends on that of the other, but when this measure is very low, these two features are considered to be independent.

Deep neural network architectures (i.e., neural networks with more than one hidden layer) are highly popular in the machine learning community due to their high capability for modeling data. However, having more layers means that more parameters are required to be tuned during the training phase. Therefore, there is a risk of overfitting to the training data as well as the network falling into a local minimum. Additionally, tuning more parameters brings computational issues, such as memory limitation and increased training time.

Input: D: Data, L: #Layers

Output: F: Feature Vector

Definition: W: Encoding weights, b: Encoding bias

Step 1: W': Decoding weights, b': decoding bias, q(.): corrupting function;

Step 2: l←1;

Step 3: $\widetilde{D}_1 \leftarrow D$;

Step 4: while $l \leq L$ do

Step 4.1: Initialize new layer with W_l, b_l, W_l', b_l' ;

Step 4.2: Correct the input: $\breve{D}_l \Leftarrow q(\widetilde{D}_l)$

Step 4.3: Train the model with W_l, b_l, W_l', b_l' with input $\breve{D}_l$ and output D;

Step 4.4: Generate the features for the next layer $\widetilde{D}_{l+1} \leftarrow f(\widetilde{D}_l | W_l, b_l)$

Step 4.5: l←l+1 ;

Step 4.6: end

Step 5: $F \longleftarrow \widetilde{D}_{L+1}$ OR concat $(\widetilde{D}_1, \dots, \widetilde{D}_{L+1})$

A way to avoid these issues is to train each layer one by one, and then stack them on top of each other whilst keeping the weights of each trained layers static. Each hidden layer represents a level of abstraction which can be used for a classification of cybercrime incident. Usually, the representation at the final layer is considered for further analysis; however, concatenating all the abstract representations and the original data vector can also be utilized (to use information in all the abstract representation). Note that, during network construction, once a layer has been trained, it receives the uncorrupted output of the previous layer. The initialization of the weights of a deep network plays a great role in avoiding local minima.

In this proposed Autoencoder model, MI representation provides a good starting point to initialize proposed Stacked Autoencoders (SA), as the most popular input representations, such as randomly and one hot ones, are non-informative. Iteratively, the autoencoder extracts the feature vector for each word from the hidden layer. This hidden layer contains the compact, compressed, and synthesized representation compared with that in the initial input layer. Afterwards, this hidden layer is used in the following level as an input layer.

SA first trains the first-level auto-encoder on the raw input to obtain parameters W_1^1, W_2^1 (weight matrix), b_1^1, b_2^1 (bias vector) and transforms the raw input into a vector which activates the hidden layer (learning and producing the primary feature vector y^1. Then, it trains the second-level auto-encoder on this vector to obtain parameters $W_1^1, W_2^1, b_1^1,$ and b_2^1. Hence, the idea is to repeat this process for subsequent levels using the hidden layer (feature vector y^k) of the previous auto-encoder level as an input for the current auto-encoder level, and so on until

y^N, with N being the number of auto-encoders in SA. Finally, all autoencoder levels must be combined to form the SA with N−1 hidden layers.

5. CLASSIFICATION TECHNIQUES

The performance of the proposed AE-FS method is evaluated using classification techniques like Random Forest (RF), K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN).

5.1 Random Forest Classification

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or the same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithms of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks. RF classifier can be described as the collection of tree structured classifiers. It is an advanced version of Bagging such that randomness is added to it [25]. Instead of splitting each node using the best split among all variables, RF splits each node using the best among a subset of predictors randomly chosen at that node.

5.2 Artificial Neural Network Classification

Among different machine learning algorithm, ANN [26] has absorbed less attention in this area but recently has attracted more attention and popularity. The key idea of ANN is to extract features from linear combination of the input data, and then models output as a nonlinear function of these features. Neural networks are usually displayed as a network diagram which involves nodes connected by links. Nodes are arranged in a layer and the architecture of common neural networks includes three layers: input layer, output layer and a hidden layer. There are two types of neural networks, feed forward and back forward. Each connection has a corresponding weight value which is estimated by minimizing a global error function in a gradient descent training process. A neuron is a simple mathematical model which outputs a value in two steps. In first step, the neuron calculates a weighted sum of its input and then obtains its output by applying an activation function to this sum. The activation function is typically a nonlinear function and it ensures that the whole network can estimate a nonlinear function which is previously learned from the input data.

5.3 K-Nearest Neighbour Classification

KNN [27] is one of the most popular instances of learning-based methods. In this method, K is the number of considered neighbours which is usually odd, and the distance to these neighbours are determined based on the standard Euclidean intervals. The underlying assumption in this method is that, all the samples are real points in n-dimensional space. In general, this algorithm is used for two purposes: to estimate the distribution density function of training data as well as to classify testing data based on the training patterns.

6. RESULT AND DISCUSSION

6.1 Dataset Description

In this research work, the student performance dataset is considered from UCI repository [20]. From the dataset link, the student performance in the Mathematic subject is considered for this research work. Table 1 depicts the dataset features used in this research work.

Table 1: Student Performance Dataset and its feature values

Attribute Name	Values
School	student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
Sex	student's sex (binary: 'F' - female or 'M' - male)
Age	student's age (numeric: from 15 to 22)
Address	student's home address type (binary: 'U' - urban or 'R' - rural)
Fam size	family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
Pstatus	parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
Medu	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
Fedu	father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
Mjob	mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
Fjob	father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
Reason	reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
Guardian	student's guardian (nominal: 'mother', 'father' or 'other')
Travel time	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
Study time	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
Failures	number of past class failures (numeric: n if 1<=n<3, else 4)
Schools up	extra educational support (binary: yes or no)
Fams up	family educational support (binary: yes or no)
Paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
Activities	extra-curricular activities (binary: yes or no)
Nursery	attended nursery school (binary: yes or no)
Higher	wants to take higher education (binary: yes or no)
Internet	Internet access at home (binary: yes or no)

Romantic	with a romantic relationship (binary: yes or no)
Famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
Free time	free time after school (numeric: from 1 - very low to 5 - very high)
Gout	going out with friends (numeric: from 1 - very low to 5 - very high)
Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
Health	current health status (numeric: from 1 - very bad to 5 - very good)
Absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20, output target) (A – 16-20, B – 11-15, C – 6-10, D – 0-5)

6.2 Performance Metrics

Table 2 depicts the performance metrics used in this research work to evaluate the performance of the proposed AE-FS method using classifiers like KNN, ANN and RF.

Table 2: Performance Metrics used in this research work

Metrics	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$
False Positive Rate	1-Specificity
Miss Rate	1-Sensitivity
False Discovery Rate	1-Precision

6.3 Number of Features obtained by Feature Selection Methods

Table 3 depicts the number of features obtained by the Proposed AE-FS method, Information Gain (IG), Symmetrical Uncertainty (SU) techniques.

Table 3: Number of Features obtained by Feature Selection Methods

Feature Selection Methods	Number of Features obtained
Original Dataset	31
IG processed dataset	29
SU processed dataset	28
Proposed AE-FS method Processed dataset	25

6.4 Performance Evaluation of the Proposed AE-FS Method

Table 4 gives the classification accuracy of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 4, it is clear that the proposed AE-FS method with ANN classification gives more accuracy when it is compared other FS methods with other classifiers.

Table 4: Classification Accuracy (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	Classification Accuracy (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	65.42	60.28	52.76
Proposed AE-FS Method	88.43	73.81	70.21
IG	75.62	63.24	56.72
SU	72.21	44.32	41.26

Table 5 gives the Sensitivity (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 5, it is clear that the proposed AE-FS method with ANN classification gives more sensitivity when it is compared other FS methods with other classifiers.

Table 5: Sensitivity (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	Sensitivity (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	64.31	59.37	52.64
Proposed AE-FS Method	79.91	71.46	68.81
IG	62.95	61.72	55.27
SU	60.59	46.64	40.32

Table 6 gives the Specificity (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 6, it is clear that the proposed AE-FS method with ANN classification gives more specificity when it is compared other FS methods with other classifiers.

Table 6: Specificity (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	Specificity (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	63.42	58.46	51.75
Proposed AE-FS Method	78.88	64.45	62.12
IG	68.23	52.27	44.86
SU	65.35	36.91	32.84

Table 7 gives the Precision (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 7, it is clear that the proposed AE-FS method with ANN classification gives increased precision when it is compared other FS methods with other classifiers.

Table 7: Precision (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	Precision (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	61.33	58.88	40.64
Proposed AE-FS Method	80.64	70.72	63.29
IG	63.49	58.46	59.57
SU	58.13	52.54	35.92

Table 8 gives the False Positive Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 8, it is clear that the proposed AE-FS method with ANN classification gives reduced FPR when it is compared other FS methods with other classifiers.

Table 8: False Positive Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	False Positive Rate (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	36.58	41.54	48.25
Proposed AE-FS Method	21.12	35.55	37.88
IG	31.77	47.73	55.14
SU	34.65	63.09	67.16

Table 9 gives the Miss Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 9, it is clear that the proposed AE-FS method with ANN classification gives reduced Miss Rate when it is compared other FS methods with other classifiers.

Table 9: Miss Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	Miss Rate (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	35.69	40.63	47.36
Proposed AE-FS Method	20.09	28.54	31.19
IG	37.05	38.28	44.73
SU	39.41	53.36	59.68

Table 10 gives the False Discovery Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification. From the table 10, it is clear that the proposed AE-FS method with ANN classification gives reduced FDR when it is compared other FS methods with other classifiers.

Table 10: False Discovery Rate (in %) of the Original dataset, Proposed AE-FS method, IG, SU processed datasets using ANN, RF and KNN classification

Feature Selection Methods	False Discovery Rate (in %) by Classification Techniques		
	ANN	RF	KNN
Original Dataset	38.67	41.12	49.36
Proposed AE-FS Method	19.36	29.28	36.71
IG	36.51	48.54	50.43
SU	41.87	51.46	64.08

7. CONCLUSION

For a productive and a good life, education is a necessity and it improves individuals' life with value and excellence. Also, education is considered a vital need for motivating self-assurance as well as providing the things are needed to partake in today's World. Throughout the years, education faced a number of challenges. Different methods of teaching and learning are suggested to increase the learning quality. For enhancing the recommendation of the student learning process by introducing the classification of the student performance into two category like slow learners and fast learner. To improve that classification accuracy, in this research paper an Auto-Encoder based Feature Selection method is proposed with the help of Mutual Information based Filter feature selection and Auto-Encoder. The performance of the proposed AE-FS method is evaluated using classification techniques like RF, ANN and KNN. This proposed AE-FS method increases the classification accuracy, specificity, sensitivity and precision with ANN than other classifiers. It also reduced the error rates like FPR, Miss Rate and FDR.

REFERENCES

[1] Amazona, Mayreen V., and Alexander A. Hernandez. "Modelling student performance using data mining techniques: Inputs for academic program development." Proceedings of the 2019 5th International Conference on Computing and Data Engineering. 2019.
[2] Shah, Mehil B., Maheeka Kaistha, and Yogesh Gupta. "Student Performance Assessment and Prediction System using Machine Learning." 2019 4th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2019.
[3] Sokkhey, Phauk, and Takeo Okazaki. "Comparative Study of Prediction Models on High School Student Performance in Mathematics." 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, 2019.
[4] Salal, Y. K., S. M. Abdullaev, and Mukesh Kumar. "Educational data mining: Student performance prediction in academic." IJ of Engineering and Advanced Tech 8.4C (2019): 54-59.
[5] El Gamal, A. F. "An educational data mining model for predicting student performance in programming course." International Journal of Computer Applications 70.17 (2013): 22-28.
[6] Saa, Amjad Abu. "Educational data mining & students' performance prediction." International Journal of Advanced Computer Science and Applications 7.5 (2016): 212-220.
[7] Sekeroglu, Boran, Kamil Dimililer, and Kubra Tuncal. "Student performance prediction and classification using machine learning algorithms." Proceedings of the 2019 8th International Conference on Educational and Information Technology. 2019.
[8] Pojon, Murat. Using machine learning to predict student performance. MS thesis. 2017.
[9] Xu, Jie, Kyeong Ho Moon, and Mihaela Van Der Schaar. "A machine learning approach for tracking and predicting student performance in degree programs." IEEE Journal of Selected Topics in Signal Processing 11.5 (2017): 742-753.
[10] Jain, Abhinav, and Shano Solanki. "An efficient approach for multiclass student performance prediction based upon machine learning." 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019.

Related Works

[11] Turabieh, Hamza, et al. "Enhanced harris hawks optimization as a feature selection for the prediction of student performance." Computing (2021): 1-22.
[12] Turabieh, Hamza. "Hybrid machine learning classifiers to predict student performance." 2019 2nd international conference on new trends in computing sciences (ICTCS). IEEE, 2019.
[13] Gajwani, Juhi, and Pinaki Chakraborty. "Students' Performance Prediction Using Feature Selection and Supervised Machine Learning Algorithms." International Conference on Innovative Computing and Communications. Springer, Singapore, 2021.

[14] Sultana, Jabeen, M. Usha, and M. A. H. Farquad. "Student's performance prediction using deep learning and data mining methods." Int. J. Recent Technol. Eng.(IJRTE)(1S4) (2019): 1018-1021.

[15] Sokkhey, Phauk, et al. "Multi-models of educational data mining for predicting student performance in mathematics: A case study on high schools in Cambodia." IEIE Transactions on Smart Processing and Computing 9.3 (2020): 217-229.

[16] Nagy, Marcell, and Roland Molontay. "Predicting dropout in higher education based on secondary school performance." 2018 IEEE 22nd international conference on intelligent engineering systems (INES). IEEE, 2018.

[17] Naicker, Nalindren, Timothy Adeliyi, and Jeanette Wing. "Linear support vector machines for prediction of student performance in school-based education." Mathematical Problems in Engineering 2020 (2020).

[18] Arif, Md, et al. "An Improved Prediction System of Students' Performance Using Classification model and Feature Selection Algorithm." International Journal of Advances in Soft Computing & Its Applications 13.1 (2021).

[19] Hidalgo, Ángel Casado, Pablo Moreno Ger, and Luis De La Fuente Valentín. "Using Meta-Learning to predict student performance in virtual learning environments." Applied Intelligence (2021): 1-14.

[20] https://archive.ics.uci.edu/ml/datasets/student+performance