International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2024): 6.623

An Overview of Early Prediction and Classification Models for the Performance Analysis of Degree Students

Gurav Bharati[1], Dr. Shubhangi Sapkal[2]

[1]Department of Comp. Sci. & Engg, Research Schlor, NIELIT, Chhatrapati Sambhajinagar, Maharashtra, India
Email: *bharatigurav2008[at]gmail.com*

[2]MCA Department, Assistant. Professor, Government College of Engineering, Chhatrapati Sambhajinagar, Maharashtra, India
Email: *shubhangisapkal24[at]gmail.com*

Abstract: ***The students who are working to earn a bachelor's degree at a college or university are known as degree students. The success of the degree students plays an essential role in enhancing the values of the college; also, the performance of the institution is represented by the student's success rate. Hence, in the education field, significant attention has been gained by the prediction of degree student's Academic Performance (AP). The students' performances are assessed by formal exams, quizzes, research projects, case studies, and homework and the results are represented in a Grade Point Average (GPA) format. The performances of the students are influenced by students' e - learning activity, family support, student internal assessment grade, previous assessment grade, low entry grades, accommodation, and GPA. For providing adequate support to enhance the student's performance, an effective as well as early prediction of the student's performance is required. To forecast the students' performances in the academic period, several prediction and classification models were developed by the prevailing frameworks. Hence, this review analyzes the developed prediction and classification models utilized for the early performance prediction of the students. The evaluation of the developed models exhibited the Machine Learning (ML) and Deep Learning (DL) methods' superiority in performance prediction by attaining effective prediction accuracy, precision, and recall.***

Keywords: Student's academic performance, Prediction model, Deep learning, Data Mining (DM), Review, Classification algorithms, Machine learning, and Learning Analytics (LA)

1. Introduction

The degree students are the undergraduate students who pursue the first level of higher education at a college or university to earn a bachelor's degree. The reputation and pride of the college or university are grounded on the academic achievements attained by the students. The extent to which a student, teacher, or institute has achieved their short - term or long - term educational goals is the academic achievements or AP [1]. AP, which covers a wide range of academic fields, is the progress made toward the goal of acquiring educational knowledge, skills, and materials [2]. Academic awards, GPA, sporting achievements, participation in competitions, leading a club or committee, and honors are some academic achievements of the degree students [3]. Currently, a new platform named Massive Open Online Course (MOOC) has emerged to enhance the learning performance of students in distance education. An e - learning platform that encloses online teaching and course delivery is named MOOC, which is less expensive when compared with traditional modes of learning [4]. The performance of the degree students relies on their studying habits, time management skills, health condition, family relationship, learning ability, teacher - student relationship, and lifestyle [5]. Since these factors are dynamic, the performance of students is not always stable [6]. Therefore, the prediction of students' performances is very essential, and it aids educators in monitoring the student's progress as both pass and fail. By predicting the students' performance, teachers can help prevent students from dropping out before final examinations and detect the students who require additional support and motivation in their studies [7].

Recently, a chief role has been played by Learning Analytics (LA) in enhancing the educational system by concentrating on various perspectives like student perspective, teacher perspective, and administrative perspective [8]. Data Mining (DM) and ML techniques are the common LA techniques employed in the educational sector. In the educational field, DM techniques play a vital role. A discipline that utilizes DM mechanisms for predicting the AP of the degree students is called Educational Data Mining (EDM) [9]. Hidden patterns are discovered by ML techniques in EDM, which then explore useful information from educational settings [10]. Figure 1 defines the process of predicting the students' AP.

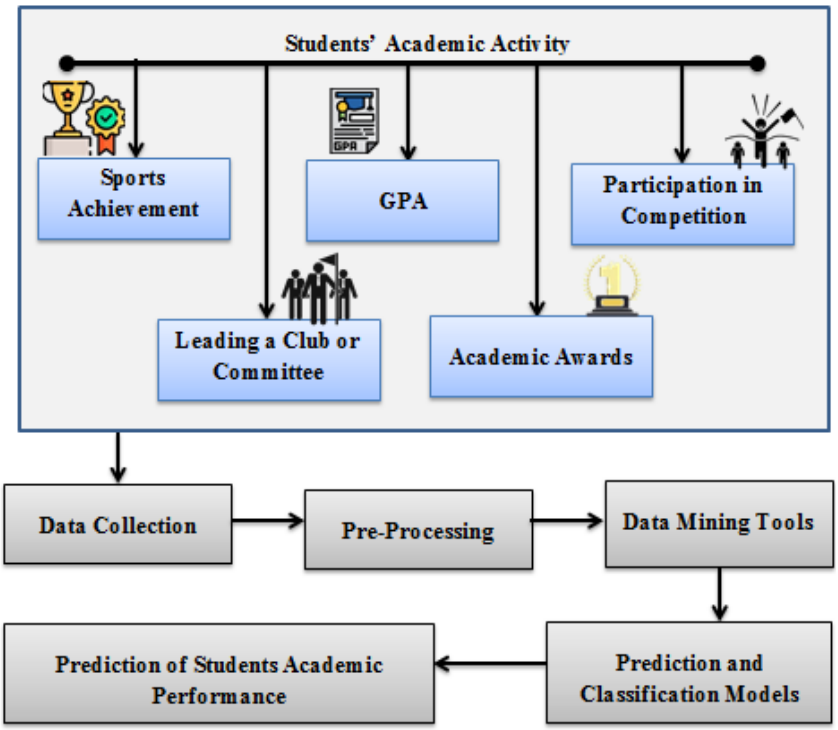

Figure 1: Process of predicting the performance of degree students

The students' AP is predicted and classified utilizing several prevailing approaches, namely Support Vector Machine (SVM), Logistic Regression (LR), Extreme Learning Machine (ELM), K - Nearest Neighbors (KNN), DM, Naive Bayes (NB), Decision Trees (DT), and Artificial Neural Network (ANN) and ensemble techniques, namely Adaptive Boosting (AB), Random Forest (RF), and Bagging. Nevertheless, some of the challenges of the developed prediction and classification models for the student's performance analysis are data availability, model interpretability, quality of the data, and model complexity. Thus, this review assesses various prediction and classification models utilized in the forecasting of degree students' future performance and briefly describes the pros and cons of the developed models to indicate the models' effectiveness. Figure 2 exemplifies the architecture of the review.

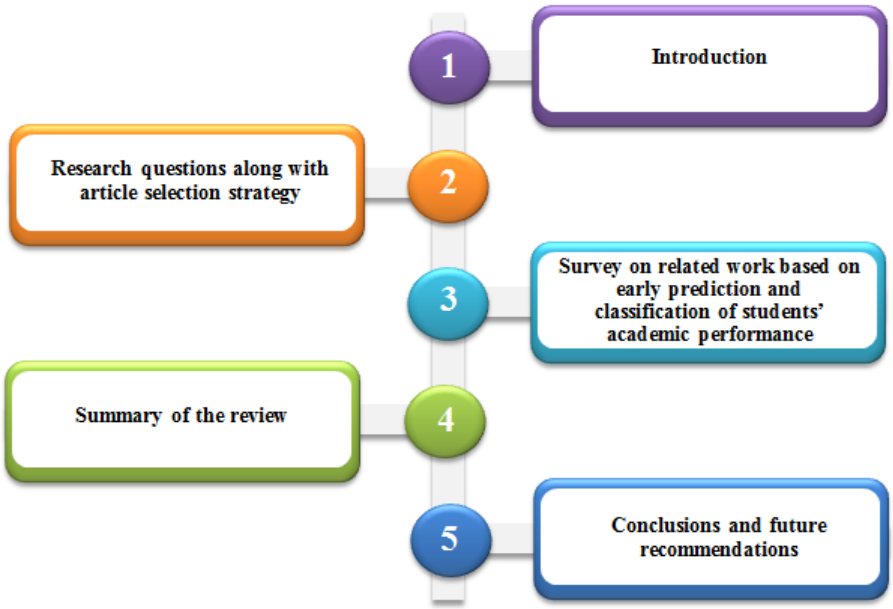

Figure 2: Architecture of the review

2. Research Questions along with Article Selection Strategy

In this analysis, the overview of early prediction and classification models utilized for predicting the performance of degree students is carried out. The analysis is processed in a way grounded on responding to the research questions, which are the reflection of the review objectives. For the analysis, the articles of the prevailing works are utilized.

Here, the selection criteria for the utilized articles are also discussed.

2.1 Research questions

- How is the degree student's AP defined? Describe different data used to predict the performance of degree students.
- State some databases used for the evaluation of the student's performance prediction and classification models.
- How effective are the LA techniques in students' performance prediction?
- Describe different DM mechanisms utilized to predict students' performance.
- How the students' performances are predicted earlier by utilizing various ML systems?
- What are the different DL models employed for the classification of the student's AP?
- Which is an effective model to predict and classify the performance of degree students earlier?

2.2 Article selection strategy

The articles utilized in the review are selected grounded on some strategies and it is described under the inclusion as well as exclusion criteria of the review. The inclusion criteria for the article selection are described as follows. The articles for the survey are chosen centered on the keywords associated with the objective of the review. The records are collected from standard search engines, such as Elsevier, Springer, IEEE, and other article sources in Google Scholar. For the analysis, the articles published after 2016 are only selected. The articles written only in the English language are utilized. For the survey, the records based on the objective of the review are selected. Then, the exclusion criteria of the review are as follows. For the analysis, the articles other than English language are not selected. The articles published before 2016 are neglected, and the articles without the objective of performance prediction of degree students are also omitted. By selecting 114 articles from reputed articles sources like Elsevier, IEEE, Springer, and other journals in Google Scholar, the article selection process is initiated. First, 114 articles were analyzed for similarity and 14 records were removed for duplication. Thereafter, 100 articles were processed for the English articles and 3 records were omitted rather than the English written articles. Also, 97 records were assessed for the publication year. Here, 13 articles were neglected for publication before 2016. Afterward, 4 records are neglected as they are books, letters, and hypothesis articles. Lastly, in this review, 80 articles were included. Figure 3 graphically displays the article selection strategy for the review.

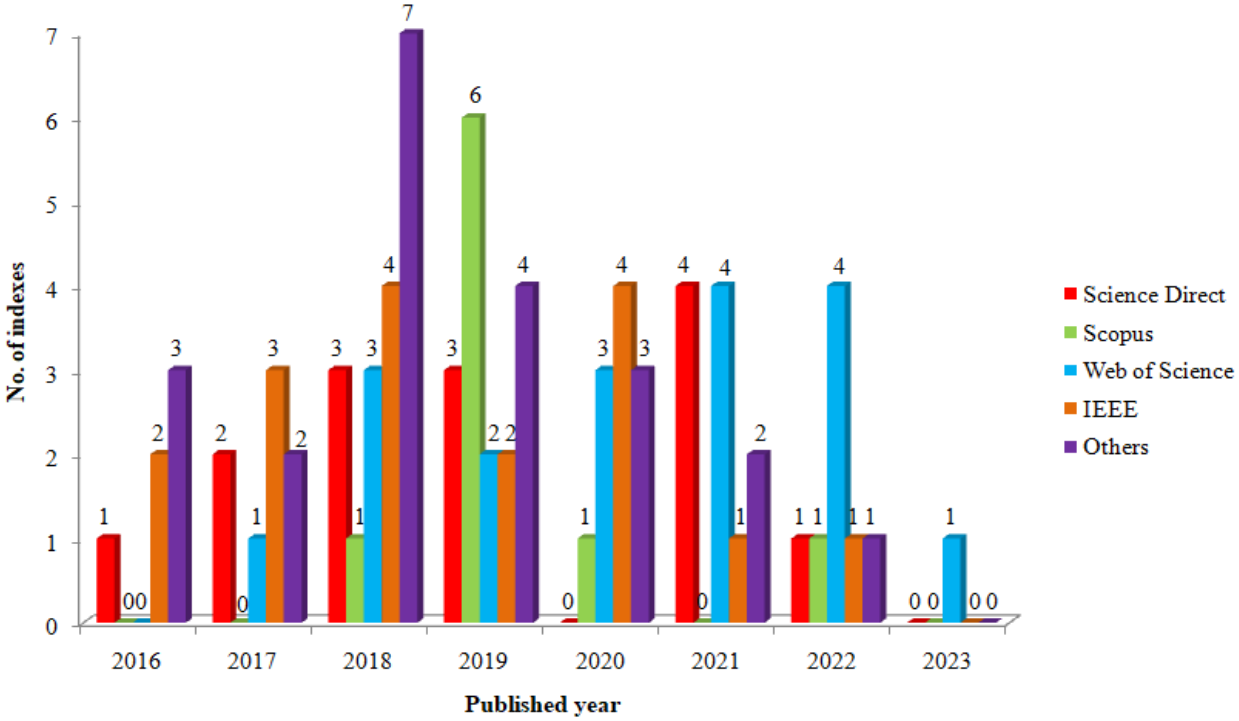

Figure 3: Graphical representation of article selection strategy

3. Literature Survey

For the early prediction and classification of degree students' performances in the academic period, several techniques were developed in the prevailing works. In section 3.1, different data utilized to predict the performance of degree students are stated; different datasets utilized for the evaluation of the student's performance prediction and classification models are presented in section 3.2; in section 3.3, the LA in students' performance prediction is indicated; different DM mechanisms in predicting students' performance are given in section 3.4; ML approaches in students' performance prediction are described in section 3.5; in section 3.6, different DL models employed for the classification of the student's AP are stated; lastly, the performance evaluation of different prediction and classification models for the degree students is presented in section 3.7.

3.1 Overview of degree students' academic performance and data used for predicting the students' performance

The evaluation of the student's achievements across various academic activities is defined as the AP of higher education students [11]. The performances of the students based on their grades, results from formal tests, quizzes, sports involvement, and classroom behaviors are measured by the educators and instructors [12]. Several past information of the students were utilized as input to measure the future performance of the students.

Tsung - Yen Yang *et al.* [13] established a behavioral - based prediction system for the grade evolution of students in MOOCs utilizing time neural networks. The framework employed the richer form of data gathered from every single student through their behaviors in video lectures and watching click streams. With two MOOC datasets, the experimental evaluation was performed. Also, the evaluation results stated that this model effectively aided educators in detecting struggling as well as advanced students earlier. However, the efficiency of the developed model was significantly reduced owing to the input features' low correlation.

Sumyea Helal *et al.* [14] elucidated the students' AP prediction model by considering student heterogeneity data like enrolment data and course activity data. The developed model was grounded on the observation of the students' study modes, learning motivations, and different socio - demographic features. As per the experimental validation, considering both enrolments as well as course activity data was effective in detecting vulnerable students more precisely. Nevertheless, this model did not concentrate on the features of the learning module for the student's participation in the learning activity. The different input data utilized by the student's AP prediction models are exemplified in Table 1.

Table 1: Analyzing different input data used for predicting students' academic performances

References	Input data	Purpose	No. of Students	Challenges
Ilie Gligorea *et al.* [15]	Engagement of students in online learning	To predict the students' AP	32, 593	The hidden features in the students' data were not analyzed
Md. Saddam Hossain Mukta *et al.* [16]	Textual contents of students' Facebook news feeds	AP evaluation	357	The developed model limited the ability to make causal inferences
Huaxiu Yao *et al.* [17]	Longitudinal behavioral data	Predicting AP for College Students	6597	Some important factors, namely intellectual level, psychological factors, and sports participation were not considered
Samara Ruiz *et al.* [18]	Students' emotional and attendance data	To predict students' outcomes from emotional responses in the classroom as well as attendance	226	The students' at - risk prediction was not accurate
Zhuojia Xu *et al.* [19]	Learning behavior	Student Performance Prediction	127	Prediction was unstable in the earlier stage

3.2 Different datasets utilized for the evaluation of the students' performance prediction and classification models

The datasets were valuable sources that helped to enhance the students' AP by providing the past learning activity data of the degree students [20] [21]. The datasets were a rich collection of features, which were required for the predictive models' development [22]. The features of the datasets provide the factors that lead to success and failure in the students' academics [23]. Here, the different datasets employed in the predictive models of students' AP were discussed.

Abdallah Moubayed *et al.* [24] established a student engagement level detection model in an e - learning environment. K - Means clustering was utilized by this framework for clustering the students grounded on twelve engagement metrics under interaction - related as well as effort - related categories. This model's experimental evaluation was performed with the University of Western Ontario (UWO) dataset. Therefore, the students with low engagement levels were effectively identified by this model. Yet, the students' demographics were not included in this model.

Jie Xu *et al.* [25] introduced an ML mechanism to predict student's performance in degree programs. A bi - layered structure and a data - driven system were comprised in this framework for making predictions on students' performance. The developed model was analyzed by the University of California, Los Angeles (UCLA). The developed model's superiority over the benchmark approaches was proved by the analysis. However, this model had an increased complexity rate owing to large feature vectors.

Some of the datasets used for the evaluation of the prediction models of students' higher education performances are elucidated in Table 2. Technical Institutes in the State of Uttar Pradesh (TIUP) dataset, Experience Application Programming Interface (XAPI) dataset, Open University Learning Analytics Dataset (OULAD), ASSISTments dataset, Andes dataset, simulated dataset, and Saudi Public University (SPU) dataset are the datasets utilized.

Table 2: Describing different datasets used in predicting students' academic performances

Authors	Techniques	Datasets	No. of records	Findings	Demerits
Deepti Aggarwal et al. [26]	Ensemble Learning	TIUP	6807	F1 score - 93.8%	No correlation
Samuel - Soma M. Ajibade et al. [27]	Ensemble Methods	XAPI	500	Accuracy - 92.3%	High noise occurrence
Qi Liu et al. [28]	Exercise - Enhanced Recurrent Neural Network and Bidirectional Long Short Term Memory (Bi - LSTM)	Large - scale real - world dataset	84909	Accuracy - 84%	High complexity
Bindhia K. Francis and Suvanam Sasidhar Babu [29]	EDM	Real - time student dataset	-	Accuracy - 75.47%	Overfitting issue
Martin Hlosta et al. [30]	Ouroboros approach	OULAD	2500	Recall - 97.92%	Biased data
Amal Asselman et al. [31]	XGBoost	ASSISTments, Andes, and Simulated datasets	462	Average accuracy - 75.24%	Ineffective for courses with no history
Hanan Abdullah Mengash [32]	ANN	SPU	2039	Accuracy - 79%	Lack of transparency

Dabiah Alboaneen *et al.* [33] propounded a web - centered prediction model for students' AP with educational institution datasets. By utilizing academic as well as demographic factors, the students who were at risk of failure were identified by this model. As per the outcomes, the developed model predicted the students' AP with minimum Mean Absolute Percentage Error (MAPE). Nevertheless, this model produced effective outcomes only for female students.

3.3 Learning analytics in students' performance prediction

The LA was employed to discover the patterns and relationships among the data utilized in the performance prediction of the degree students by employing many tools and techniques. LA is the measurement, analysis, reporting, and collection of data about the learners for predicting their ability and needs in learning. A multidisciplinary field that involves Artificial Intelligence (AI), ML, DL, statistics, information retrieval, and visualization is named LA [34]. The LA approaches in students' performance prediction process are displayed in Table 3.

Table 3: Evaluating learning analytics in students' performance prediction

References	Objectives	Techniques	Parameters	Achievements			Limitations
				Precision (%)	Accuracy (%)	Recall (%)	
Elvira Popescu and Florin Leon [35]	To predict AP centered on Learner Traces in a Social Learning Environment	Large Margin Nearest Neighbor Regression	Web Applications Design course data	-	85	-	Did not adapt distance metric
Mustafa Yağcı [36]	ML - based approach for predicting the final exam grades of undergraduate students	EDM	Midterm exam grades, Department data, and Faculty data	75.2	75	74.6	Limited visualization data
Raza Hasan et al. [37]	To predict student performance in higher educational institutions utilizing video LA	EDM	Student information system, learning management system, and mobile application data	-	88.3	-	Data tracking was a complex issue
Farshid Marbouti et al. [38]	Early prediction of at - risk students in a course utilizing standards - based grading	NB	Achievement of learning objectives and course grades data	-	92.6	-	The evaluation process was complex
Omiros Iatrellis et al. [39]	Two - phase ML technique to predict student outcomes	Unsupervised and supervised learning techniques	Behavioral data	79	80	84	Efficiency depends on high - quality data

Owen H. T. Lu *et al.* [40] employed an early prediction system for students' AP in blended learning. To predict the students' final outcomes, the framework utilized LA and big data approaches. For the prediction, the course, quiz scores, video - viewing behavior, homework, out - of - class practice behaviors, and after - school tutoring were analyzed. The analysis results exposed that this model had higher predictive performance in contrast to the state - of - art models. However, owing to processing large amounts of records, the developed model had considerably higher misclassification issues than the other models.

David Azcona *et al.* [41] proffered a detection model for finding the students at risk in computer programming classes. The predictive analytics was implemented with LA from students' digital footprints from blended classroom settings. According to the outcomes, the developed technique efficiently helped in reducing the gap betwixt the

low and high - performing students. Yet, features from diverse data sources were only taken, which increased the possibility of prediction errors.

3.4 Different data mining techniques for predicting students' performance

Enhancing education quality to a standard level by maximizing academic results and mitigating the students' failure rate was the primary goal of the educational system [42]. The rapid increase in educational data had difficulty in processing through traditional methods. Hence, DM has been utilized for the past few years because of its ability to analyze large amounts of data quickly [43]. EDM is utilized for predicting students' AP earlier since it is an effective discipline of DM in the educational field [44]. EDM transforms the raw information of the students from large educational databases to useful information, which is further utilized to understand the learning conditions of the students [45].

Ashraf Abazeed and Moaiad Khder /46] introduced a classification as well as prediction model with DM methods for students' performance at the university level. To find the hidden information in the student's records, association rules were utilized in this framework. This model aided the students in determining their direction of study and improving their learning methodologies whenever it was necessary. As per the analysis, the developed model effectively predicted the students' who needed attention and who were deviating from their studies. However, this model was ineffective with unstructured data in the students' databases.

Vera L. Miguéis *et al.* [47] developed a predictive modeling framework supported by DM techniques for the early segmentation of students grounded on their APs. Here, the students were segmented centered on the difference between their performance at the beginning of the degree program and the predicted performance level of the students. According to the outcomes, this technique could predict students' performance levels earlier with better prediction accuracy. Nevertheless, inaccurate estimation of relationships between variables was possible in this model as it was a sample - biased model. Table 4 exemplifies the different DM mechanisms in predicting degree students' performance.

Table 4: DM techniques in predicting students' performance

Authors	DM techniques	No. of instances	Attributes/ Features	Classification Accuracy (%)	Limitations
Hilal Almarabeh [48]	NB, BN, ID3, J48 and NN	225	10	92	Surveying timing was high
Elaf Abu Amrieh et al. [49]	Bagging, boosting, and RF	500	16	80	Inapplicable with smaller data
Olugbenga Wilson Adejo and Thomas Connolly [50]	DT, ANN, and SVM	505	30	81.67	High classification error
Y. K. Saheed et al. [51]	ID3, C4.5, and CART	234	-	98.3	Complex data
Concepción Burgos et al. [52]	LR	104	-	85	Gathered additional irrelevant information
Sadiq Hussain et al. [53]	J48, PART, RF, and BN	300	24	77.91	Comparatively poor performance
Ahmed Mueen et al. [54]	NB, NN, and DT	60	38	86	Require a large amount of data

The different DM techniques used for predicting and classifying the APs of the degree students are exemplified in Table 4. NB, Bayesian Network (BN), DT, ANN, SVM, Iterative Dichotomiser 3 (ID3), J48, Classification and Regression tree (CART), LR, PART, RF, and Neural Network (NN) are some of the DM techniques employed.

Ralph Olusola Aluko *et al.* [55] established a framework based on predicting the AP of architecture students using DM techniques. Here, LR and SVM were utilized for the performance prediction by utilizing the information from prior academic achievements. As per the analysis, the developed model performed better than the other conventional models in the students' performance prediction. However, the system was not suitable for the student's performance prediction using categorical dependent variables.

3.5 Machine learning techniques - based higher education students' academic performance prediction

Students' APs were predicted earlier to reduce the dropout rates, assist with their studies, and improve their grades [56]. In the early prediction of the student's academic activities, the ML technique obtained robust performance. ML approaches render timely and actionable information regarding the degree students' performances [57] [58]. Through early identification of students' needs, ML can transform the overall learning experience of the students and the educational institutions [59]. The different ML techniques deployed for the early performance prediction of higher education students are depicted in Table 5,

Table 5: Analyzing different ML techniques in academic performance prediction of higher education students

References	ML techniques	Sources	Dataset description			Performance metrics - Values (%)	Challenges
			Total data	Male	Female		
E. T. Lau *et al.* [60]	ANN	University Q	1000	275	810	Accuracy - 84.8 Precision - 86.3 Sensitivity - 94.8	A high false negative rate
Okereke GE *et al.* [61]	DT	University of Nigeria	1600	921	679	Accuracy - 92.27 Precision - 58.33	Misclassification owing to overfitting issue

Cameron C. Gray and Dave Perkins [62]	Candidate ML	Bangor Institution	4970	-	-	Accuracy - 97 Precision - 87.8 Recall - 91.4	Data overlapping issue
Hassan Zeineddine *et al.* [63]	Auto ML	United Arab Emirates	1491	1283	208	Accuracy - 75.9 Kappa value - 0.5	Doesn't reveal the hidden information in the data
Mushtaq Hussain *et al.* [64]	DT, CART, and NB	Virtual learning environment	305	-	-	Accuracy - 84.89 Kappa value - 0.54 Recall - 91.97	Error - prone model
Kwok Tai Chui *et al.* [65]	Reduced Training Vector - based SVM (RTVSVM)	OULAD	32593	19229	13364	Accuracy - 93.8 Sensitivity - 94 Specificity - 93.6	Effective only for large data
Muhammad Imran *et al.* [66]	DT	UCI Machine Learning Repository	1044	-	-	Accuracy - 95.78	Unstable model

Nick Z. Zacharis [67] propounded the students' AP prediction model in blended learning and ANN. The Moodle server that stored the students' data processed this framework. Next, in a blended learning course environment, a Multi - Layer Perceptron NN was trained for the students' performance prediction. The developed system's superiority with efficient classification accuracy was stated by its analysis. Nevertheless, factors that affect the students' performance were not concentrated.

Abdullah Alshanqiti and Abdallah Namoun [68] explored a model to predict the students' performance by utilizing Hybrid Regression (HR) and Multi - Label Classification (MLC). The prediction accuracy of the students' AP was optimized by the HR; then, the qualitative values for the influenced factors of the students' performance were predicted by MLC. According to the outcomes, the developed model attained superior results in detecting the factors that influenced the students' AP. But, to train the model, the developed model required a larger time and resources.

3.6 Different deep learning models employed for the classification of the student's academic performance

An essential part of the education system is the early prediction of the student's performance. By using several hidden layer processes, the DL approaches efficiently predicted the performances of the degree students earlier for identifying the minute details to improve the learning of the predictive models. To produce accurate predictions, DL models can significantly recognize the complex patterns of data [69].

Byung - Hak Kim *et al.* [70] presented a DL - centric GridNet for predicting students' performance. The GridNet was centered on the Long Short Term Memory (LSTM) architecture to predict the students' performance and demonstrated with Udacity's Nanodegree programs. As per the analysis outcomes, the developed system achieved robust results in predicting the students' performances with students' real - time data. However, it failed to capture long - term correlation among the data, and the developed model's performance was affected.

Alberto Rivas *et al.* [71] established an AP prediction model for the students in VLE centered on ANN analysis. Various tree - centric and ANN models were included; thus, the key factors that affect the students' performances were efficiently identified. In predicting the students' performance earlier, this model's experimental analysis displayed significant performance. However, this model's prediction efficiency decreased significantly after certain epochs. Table 6 indicates the different types of DL techniques utilized for the prediction of degree students' performances.

Table 6: Analyzing different DL approaches based on their evaluation outcomes in students' performance prediction

Authors	Objective	Methodology	Purpose of DL	Accuracy (%)	Disadvantages
Aya Nabil *et al.* [72]	To predict students' AP centered on course grades	Deep Neural Network (DNN)	Predict students' AP	89	It did not have any effective models for analyzing the behavioral patterns from the students' data
Fangyao Xu and Shaojie Qu [73]	Performance prediction by DM of students' learning consumption behavior pattern	GNN	Classification of Students' AP	84.86	Few samples were used for training, which leads to misclassification
F. Giannakas *et al.* [74]	Early prediction of team - centric AP	DNN	Binary classification for team performance	82.39	Due to the early triggering of stopping callback, an efficient validation rate was not attained
Saeed-Ul Hassan *et al.* [75]	Virtual learning environment to predict students' withdrawal	LSTM	Predict early withdrawal of students	97.25	Vanishing gradient problem
Naif Radi Aljohani *et al.* [76]	For predicting at - risk students utilizing click stream data in the virtual learning environment	LSTM	Students' performance prediction	95.23	Applicable to find students' performances in a single course only
Nida Aslam *et al.* [77]	Early student's performance prediction	DNN	Predicting students' AP	96.4	Validated with a small dataset

Bashir Khan Yousafzai *et al.* [78]	Student AP prediction	BiLSTM	Grade prediction	90.16	Noisy and large data affected prediction efficiency
Hajra Waheed *et al.* [79]	To predict the students' AP from VLE big data	Deep ANN	Predict at - risk students	93	Poor performance for continuous data

Xizhe Wang *et al.* [80] explored a fine - grained learning performance prediction of students enrolled in online courses via an adaptive sparse self - attention network. The missing values were generated by the employed attention network, which predicted the fine - grained performance simultaneously. This model's analysis was demonstrated with the real - time dataset from various e - learning platforms. Hence, the analysis proved the developed model's importance. But, centered on the gradual lengthening, this model's efficacy declined.

3.7 Overall evaluation based on different prediction and classification models for degree students' performances.

In the overall evaluation of prevailing models, an efficient prediction and classification model to forecast degree students' performance was estimated. Table 7 discusses and tabulates the achievements of the existing prediction and classification models, which were used for the prediction of the student's AP.

Table 7: Overall evaluation of the existing models

Techniques	Accuracy (%)
NB [38]	92.6
SVM [65]	93.8
DT [66]	95.78
LSTM [75]	97.25
DNN [77]	96.4

In Table 7, the overall evaluation of the existing prediction and classification models used in the prevailing works for the early prediction of the students' APs centered on the student's past learning factors is indicated. The evaluation was carried out centered on diverse accuracy rates achieved by the prevailing models to predict the performance of degree students, and different techniques selected for the evaluation were NB [38], SVM [65], DT [66], LSTM [75], and DNN [77]. Next, the analysis signifies that to predict the students' AP, the LSTM [75] attained the highest accuracy rate of 97.25%. Then, the DNN [77] attained the second - highest accuracy rate of 96.4% for the students' performance prediction. After that, in forecasting the AP of the students, the DT [66] scored the third - highest accuracy rate (95.78%). Lastly, for the early AP of the degree students, the SVM [65] and NB [38] models attained 93.8% and 92.6% accuracy, respectively. Next, figure 4 graphically presents the analyzed prevailing models.

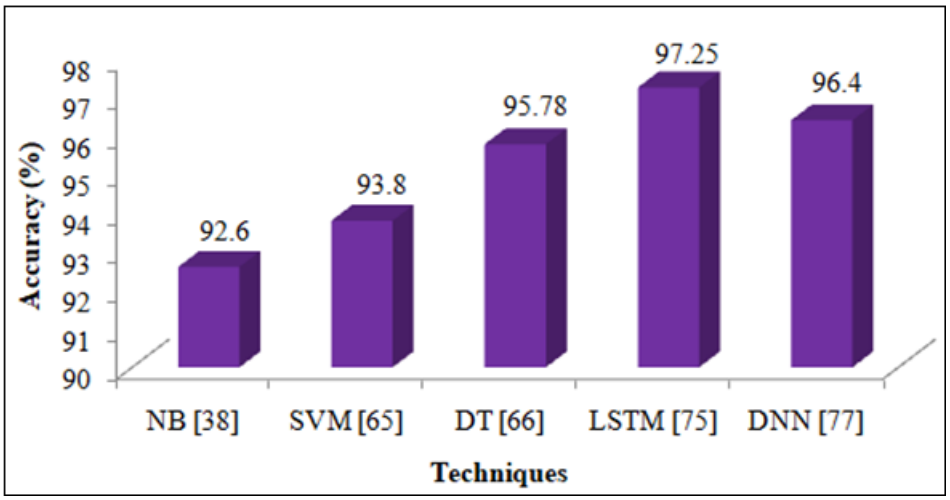

Figure 4: Comparison evaluation of the existing models

The comparison evaluation of the prevailing models centered on the accuracy achieved to predict the AP of degree students is described in Figure 4. The figure displays that for the early prediction of the student's AP, the LSTM [75] achieved the highest accuracy rate (97.25%). Thus, a robust prediction model to forecast the students' AP is obtained from the analysis.

4. Summary of the Review

The values of the educational institutions were enhanced by the academic achievements of the degree students. Thus, in the education field, the forecasting of students' AP gained attention widely. Academic achievements are the primary goal of a student, teacher, or institution. Academic awards, GPA, sporting achievements, participation in competitions, leading a club or committee, and honors are some academic achievements of higher education. To enhance their knowledge through virtual learning, students can enroll in several online platforms, such as MOOCs. However, some significant factors, namely studying habits, time management skills, health conditions, family relationships, learning ability, teacher - student relationship, and lifestyle affect the degree students' learning. The performance of the students is influenced by such unstable factors. Hence, to

enhance students' performance, reduce the weak students' dropout rate, and identify the students who require additional support for their studies, an early forecasting of degree students' performance is necessary. For the early performance prediction of higher education students, this survey significantly overviewed the various prediction and classification models employed in the prevailing works. The developed models' achievements, advantages, and disadvantages are examined and discussed under 7 research questions. To answer the research questions, the survey is divided into numerous sections. Section 3.1 presents the overview of students' performance prediction and various input data utilized to analyze students' AP. Next, section 3.2 discusses various educational databases used for the performance prediction models evaluation. After that, section 3.3 indicates the purpose of LA in the students' performance prediction. The instructors are provided with valuable criteria about the learning patterns by LA, which helps the students to identify their weak areas in learning that have to be developed. Section 3.4 depicts the DM approaches' next advantages, which states that the DM techniques aid in increasing the quality of education via efficient prediction results. Subsequently, sections 3.5 and 3.6 discussed the different ML and DL approaches deployed for the early prediction of the degree students' AP, respectively. Lastly, section 3.7 presents the overall evaluation of the existing prediction and classification models in the early forecasting of degree students' performance. As per the analysis, the LSTM [75] achieved the highest prediction accuracy rate (97.25%), followed by DNN [77] with the second highest accuracy rate (96.4%). Then, the DT [66] scored the third - highest accuracy rate (95.78%), whereas SVM [65] had an accuracy rate of 93.8% in forecasting the students' AP. Lastly, for the early AP of the degree students, the NB [38] model attained 92.6% accuracy. Hence, the efficient techniques for the early prediction of degree students' AP are evaluated and their robustness is proved in the overall analysis.

5. Conclusion

This review studied the early prediction of the degree student's AP by utilizing various approaches. For improving the students' performance and reducing their failure rate, early prediction is performed. In this survey, different LA, DM, ML, and DL models used for the prediction of degree students' AP were analyzed briefly. As per the experimental evaluation of the developed models by utilizing various educational databases, the developed model achieved 97.25% accuracy, 97.92% recall, and 79% precision rates in the early prediction of students' AP and also acquired a minimum error of 0.178 and a true positive rate of 0.634. The overall evaluation stated that in predicting students' AP earlier, the LSTM [75] achieved efficient results. Therefore, the purpose, process, merits, and demerits of the existing models in early forecasting of degree students' AP prediction are discussed in the whole survey.

6. Future Recommendations

Through this analysis, the strengths and weaknesses of the existing models are effectively understood. In most of the developed models, the performance of the degree students was effectively predicted; however, they didn't concentrate on providing any sufficient materials and learning support for the low - performing students centered on their emotional behavior during the learning period.

References

[1] Raza Hasan, Sellappan Palaniappan, Abdul Rafiez Abdul Razif, Salman Mahmood and Kamal Uddin Sarker, "Student academic performance prediction by using decision tree algorithm", 4th International Conference on Computer and Information Sciences, 13 - 14 August, Kuala Lumpur, Malaysia, 2018.
[2] Mustafa Agaoglu, "Predicting instructor performance using data mining techniques in higher education", IEEE Access, vol.4, pp.2379 - 2387, 2016.
[3] Mushtaq Hussain, Wenhao Zhu, Wu Zhang, Syed Muhammad Raza Abidi and Sadaqat Ali, "Using machine learning to predict student difficulties from learning session data", Artificial Intelligence Review, vol.52, pp.381 - 407, 2018.
[4] Fisnik Dalipi, Ali Shariq Imran and Zenun Kastrati, "MOOC dropout prediction using machine learning techniques: review and research challenges", IEEE Global Engineering Education Conference, 17 - 20 April, Santa Cruz de Tenerife, Spain, 2018.
[5] Seyhmus Aydogdu, "Predicting student final performance using artificial neural networks in online learning environments", Education and Information Technologies, vol.25, no.3, pp.1913 - 1927, 2019.
[6] Tenzin Doleck, David John Lemay, Ram B Basnet and Paul Bazelais, "Predictive analytics in education: A comparison of deep learning frameworks", Education and Information Technologies, vol.25, pp.1951 - 1963, 2019.
[7] Ali Salah Hashim, Wid Akeel Awadh and Alaa Khalaf Hamoud. "Student Performance Prediction Model based on Supervised Machine Learning Algorithms", IOP Conference Series: Materials Science and Engineering, vol.928, no.3, pp.1 - 18, 2020.
[8] Engr. Sana Bhutto, Isma Farah Siddiqui, Qasim Ali Arain and Maleeha Anwar, "Predicting students' academic performance through supervised machine learning", International Conference on Information Science and Communication Technology, 08 - 09 February, Karachi, Pakistan.2020.
[9] Chitra Jalota and Rashmi Agrawal, "Analysis of educational data mining using classification", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, 14 - 16 February, Faridabad, India, 2019.
[10] Sein Minn, Yi Yu, Michel C Desmarais, Feida Zhu and Jill - Jenn Vie, "Deep knowledge tracing and dynamic student classification for knowledge tracing", IEEE International Conference on Data Mining, 17 - 20 November, Singapore, 2018.
[11] Ansar Siddique, Asiya Jan, Fiaz Majeed, Adel Ibrahim Qahmash, Noorulhasan Naveed Quadri and Mohammad Osman Abdul Wahab, "Predicting academic performance using an efficient model based on fusion of classifiers", Applied Sciences, vol.11, no.24, pp.1 - 19, 2021.

[12] Asmaa Elbadrawy, Agoritsa Polyzou, Zhiyun Ren, Mackenzie Sweeney, George Karypis and Huzefa Rangwala, "Predicting student performance using personalized analytics", Computer, vol.49, no.4, pp.61 - 69, 2016.
[13] Tsung - Yen Yang, Christopher G Brinton, Carlee Joe - Wong and Mung Chiang, "Behavior - based grade prediction for MOOCs via time series neural networks", IEEE Journal of Selected Topics in Signal Processing, vol.11, no.5, pp.716 - 728, 2017.
[14] Sumyea Helal, Jiuyong Li, Lin Liu, Esmaeil Ebrahimie, Shane Dawson, Duncan J. Murray and Qi Long, "Predicting academic performance by considering student heterogeneity", Knowledge - Based Systems, vol.161, pp.134 - 146, 2018.
[15] Ilie Gligorea, Muhammad Usman Yaseen, Marius Cioca, Hortensia Gorski and Romana Oancea, "An interpretable framework for an efficient analysis of students' academic performance", Sustainability, vol.14, no.14, pp.1 - 21, 2022.
[16] Md. Saddam Hossain Mukta, Salekul Islam, Swakkhar Shatabda, Mohammed Eunus Ali and Akib Zaman, "Predicting academic performance: Analysis of students' mental health condition from social media interactions", Behavioral Sciences, vol.12, no.4, pp.1 - 18, 2022.
[17] Huaxiu Yao, Defu Lian, Yi Cao, Yifan Wu and Tao Zhou, "Predicting academic performance for college students: A campus behavior perspective", ACM Transactions on Intelligent Systems and Technology, vol.10, no.3, pp.1 - 21, 2019.
[18] Samara Ruiz, Maite Urretavizcaya, Clemente Rodriguez and Isabel Fernandez Castro, "Predicting students' outcomes from emotional response in the classroom and attendance", Interactive Learning Environments, vol.28, no.1, pp.1 - 23, 2018.
[19] Zhuojia Xu, Hua Yuan and Qishan Liu, "Student performance prediction based on blended learning", IEEE Transactions on Education, vol.64, no.1, pp.66 - 73, 2020.
[20] Abdallah Moubayed, Mohammadnoor Injadat, Abdallah Shami and Hanan Lutfiyya, "Student engagement level in e - learning environment: Clustering using K - means", American Journal of Distance Education, vol.34, no.2, pp.137 - 156, 2020.
[21] Lubna Mahmoud Abu Zohair, "Prediction of Student's performance by modelling small dataset size", International Journal of Educational Technology in Higher Education, vol.16, no.1, pp.1 - 18, 2019.
[22] Yahia Baashar, Yaman Hamed, Gamal Alkawsi, Luiz Fernando Capretz, Hitham Alhussian, Ayed Alwadain and Redhwan Al - amri, "Evaluation of postgraduate academic performance using artificial intelligence models", Alexandria Engineering Journal, vol.61, pp.9867 - 9878, 2022.
[23] Nur Izzati Mohd Talib, Nazatul Aini Abd Majid and Shahnorbanun Sahran, "Identification of student behavioral patterns in higher education using k - means clustering and support vector machine", Applied Sciences, vol.13, no.5, pp.1 - 14, 2023.
[24] Gomathy Ramaswami, Teo Susnjak, Anuradha Mathrani, James Lim and Pablo Garcia, "Using educational data mining techniques to increase the prediction accuracy of student academic performance", Information and Learning Sciences, vol.120, no.7/8, pp.451 - 467, 2019.
[25] Jie Xu, 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, vol.11, no.5, pp.742 - 753, 2017.
[26] Deepti Aggarwal and Vikram Bali, "Significance of Non - Academic Parameters for Predicting Student Performance Using Ensemble Learning Techniques", International Journal of System Dynamics Applications, vol.10, no.3, pp.1 - 12, 2021.
[27] Samuel - Soma M Ajibade, Johnry Dayupay, Dai - Long Ngo - Hoang, Oluwadare Joshua Oyebode, John Michael Sasan, Nathaniel G, Gido, Almighty C Tabuena and Anthonia Adediran, "Utilization of ensemble techniques for prediction of the academic performance of students", Journal of Optoelectronics Laser, vol.41, no.6, pp.48 - 54, 2022.
[28] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su and Guoping Hu, "EKT: Exercise - aware knowledge tracing for student performance prediction", IEEE Transactions on Knowledge and Data Engineering, vol.33, no.1, pp.100 - 115, 2019.
[29] Bindhia K Francis and Suvanam Sasidhar Babu, "Predicting academic performance of students using a hybrid data mining approach", Journal of Medical Systems, vol.43, pp.1 - 15, 2019.
[30] Martin Hlosta, Zdenek Zdrahal and Jaroslav Zendulka, "Ouroboros: Early identification of at - risk students without models based on legacy data", LAK '17: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 13 - 17 March, Vancouver British Columbia, Canada, 2017.
[31] Amal Asselman, Mohamed Khaldi and Souhaib Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm", Interactive Learning Environments, vol.31, no.6, pp.1 - 20, 2021.
[32] Hanan Abdullah Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems", IEEE Access, vol.8, pp.55462 - 55470, 2020.
[33] Dabiah Alboaneen, Modhe Almelihi, Rawan Alsubaie, Raneem Alghamdi, Lama Alshehri and Renad Alharthi, "Development of a Web - Based Prediction System for Students' Academic Performance", Data, vol.7, no.2, pp.1 - 19, 2022.
[34] Mohammed Saqr, Uno Fors and Matti Tedre, "How learning analytics can early predict underachieving students in a blended medical education course", Medical teacher, vol.39, no.7, pp.757 - 767, 2017.
[35] Elvira Popescu and Florin Leon, "Predicting academic performance based on learner traces in a social learning environment", IEEE Access, vol.6, pp.72774 - 72785, 2018.
[36] Mustafa Yagci, "Educational data mining: Prediction of students' academic performance using machine learning algorithms", Smart Learning Environments, vol.9, no.1, pp.1 - 19, 2022.
[37] Raza Hasan, Sellappan Palaniappan, Salman Mahmood, Ali Abbas, Kamal Uddin Sarker and Mian

Usman Sattar, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques", Applied Sciences, vol.10, no.11, pp.1 - 20, 2020.

[38] Farshid Marbouti, Heidi A Diefes - Dux and Krishna Madhavan, "Models for early prediction of at - risk students in a course using standards - based grading", Computers & Education, vol.103, pp.1 - 15, 2016.

[39] Omiros Iatrellis, Ilias K Savvas, Panos Fitsilis and Vassilis C Gerogiannis, "A two - phase machine learning approach for predicting student outcomes", Education and Information Technologies, vol.26, pp.69 - 88, 2020.

[40] Owen H. T. Lu, Anna Y. Q. Huang, Jeff C. H. Huang, Albert J. Q. Lin, Hiroaki Ogata and Stephen J. H. Yang, "Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning", Educational Technology & Society, vol.21, no.2, pp.220–232, 2018.

[41] David Azcona, I - Han Hsiao and Alan F Smeaton, "Detecting students - at - risk in computer programming classes with learning analytics from students' digital footprints", User Modeling and User - Adapted Interaction, vol.29, pp.759 - 788, 2019.

[42] Phauk Sokkhey and Takeo Okazaki, "Developing web - based support systems for predicting poor - performing students using educational data mining techniques", International Journal of Advanced Computer Science and Applications, vol.11, no.7, pp.23 - 32, 2020.

[43] Guiyun Feng, Muwei Fan and Yu Chen, "Analysis and prediction of students' academic performance based on educational data mining", IEEE Access, vol.10, pp.19558 - 19571, 2022.

[44] Preet Kamal and Sachin Ahuja, "An ensemble - based model for prediction of academic performance of students in undergrad professional course", Journal of Engineering, Design and Technology, vol.17, no.4, pp.769 - 781, 2019.

[45] Raheela Asif, Agathe Merceron, Syed Abbas Ali and Najmi Ghani Haider, "Analyzing undergraduate students' performance using educational data mining", Computers & education, vol.113, pp.1 - 35, 2017.

[46] Ashraf Abazeed and Moaiad Khder, "A classification and prediction model for student's performance in university level", Journal of Computer Science, vol.13, no.7, pp.228 - 233, 2017.

[47] Migueis V. L., Ana Freitas, Paulo J. V Garcia and Andre Silva, "Early segmentation of students according to their academic performance: A predictive modeling approach", Decision Support Systems, vol.115, pp.1 - 42, 2018.

[48] Hilal Almarabeh, "Analysis of students' performance by using different data mining classifiers", International Journal of Modern Education and Computer Science, vol.8, pp.9 - 15, 2017.

[49] Elaf Abu Amrieh, Thair Hamtini and Ibrahim Aljarah, "Mining educational data to predict student's academic performance using ensemble methods", International Journal of Database Theory and Application, vol.9, no.8, pp.119 - 136, 2016.

[50] Olugbenga Wilson Adejo and Thomas Connolly, "Predicting student academic performance using multi - model heterogeneous ensemble approach", Journal of Applied Research in Higher Education, vol.10, no.1, pp.61 - 75, 2018.

[51] Saheed Y. K., Oladele T. O., Akanni A. O and Ibrahim W. M., "Student performance prediction based on data mining classification techniques", Nigerian Journal of Technology, vol.37, no.4, pp.1087 - 1091, 2018.

[52] Concepcion Burgos, Maria L Campanario, David de la Pena, Juan A Lara, David Lizcano and María A Martinez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout", Computers & Electrical Engineering, vol.66, pp.1 - 16, 2017.

[53] Sadiq Hussain, Neama Abdulaziz Dahan, Fadl Mutaher Ba - Alwib and Najoua Ribata, "Educational data mining and analysis of students' academic performance using WEKA", Indonesian Journal of Electrical Engineering and Computer Science, vol.9, no.2, pp.447 - 459, 2018.

[54] Ahmed Mueen, Bassam Zafar and Umar Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques", International Journal of Modern Education and Computer Science, vol.11, pp.36 - 42, 2016.

[55] Ralph Olusola Aluko, Emmanuel Itodo Daniel, Olalekan Oshodi, Clinton Ohis Aigbavboa, Abiodun Olatunji Abisuga, "Towards reliable prediction of academic performance of architecture students using data mining techniques", Journal of Engineering, Design and Technology, vol.16, no.3, pp.385 - 397, 2018.

[56] Daiva Gostautaite and Leonidas Sakalauskas, "Multi - label classification and explanation methods for students' learning style prediction and interpretation", Applied Sciences, vol.12, no.11, pp.1 - 19, 2022.

[57] Alaa Khalaf Hamoud, Ali Salah Hashim and Wid Aqeel Awadh, "Predicting student performance in higher education institutions using decision tree analysis", International Journal of Interactive Multimedia and Artificial Intelligence, vol.5, no.2, pp.26 - 31, 2018.

[58] Xing Xu, Jianzhong Wang, Hao Peng and Ruilin Wu, "Prediction of academic performance associated with Internet usage behaviors using machine learning algorithms", Computers in Human Behavior, vol.98, pp.166 - 173, 2019.

[59] Jiun - Yu Wu, Yi - Cheng Hsiao and Mei - Wen Nian, "Using supervised machine learning on large - scale online forums to classify course - related Facebook messages in predicting learning achievement within the personal learning environment", Interactive Learning Environments, vol.28, no.1, pp.65 - 80, 2018.

[60] Lau E. T, Sun L and Yang Q, "Modelling, prediction and classification of student academic performance using artifcial neural networks", SN Applied Sciences, vol.1, pp.1 - 10, 2019.

[61] Okereke G. E, Mamah C. H, Ukekwe E. C and Nwagwu H. C, "A machine learning based framework for predicting student's academic performance", Physical Science & Biophysics Journal, vol.4, no.2, pp.1 - 6, 2020.

[62] Cameron C. Gray and Dave Perkins, "Utilizing early engagement and machine learning to predict student

outcomes", Computers & Education, vol.131, pp.22 - 32, 2019.
[63] Hassan Zeineddine, Udo Braendle and Assaad Farah, "Enhancing prediction of student success: automated machine learning approach", Computers & Electrical Engineering, vol.89, pp.1 - 22, 2021.
[64] Mushtaq Hussain, Wenhao Zhu, Wu Zhang and Syed Muhammad Raza Abidi, "Student engagement predictions in an e - learning system and their impact on student course assessment scores", Computational Intelligence and Neuroscience, vol.2018, pp.1 - 21, 2018.
[65] Kwok Tai Chui, Dennis Chun Lok Fung, Miltiadis D Lytras and Tin Miu Lam, "Predicting at - risk university students in a virtual learning environment via a machine learning algorithm", Computers in Human behavior, vol.107, pp.1 - 30, 2018.
[66] Muhammad Imran, Shahzad Latif, Danish Mehmood, Muhammad Saqlain Shah, "Student academic performance prediction using supervised learning techniques", International Journal of Emerging Technologies in Learning, vol.14, no.14, pp.92 - 104, 2019.
[67] Nick Z Zacharis, "Predicting student academic performance in blended learning using artificial neural networks", International Journal of Artificial Intelligence and Applications, vol.7, no.5, pp.17 - 29, 2016.
[68] Abdullah Alshanqiti and Abdallah Namoun, "Predicting student performance and its influential factors using hybrid regression and multi - label classification", IEEE Access, vol.8, pp.203827 - 203844, 2020.
[69] Maria Tsiakmaki, Georgios Kostopoulos, Sotiris Kotsiantis and Omiros Ragos, "Transfer learning from deep neural networks for predicting student performance", Applied Sciences, vol.10, no.6, pp.1 - 12, 2020.
[70] Byung - Hak Kim, Ethan Vizitei and Varun Ganapathi, "GritNet: Student performance prediction with deep learning", arXiv, pp.1 - 5, 2018. https: //doi. org/10.48550/arXiv.1804.07405
[71] Alberto Rivas, Alfonso González - Briones, Guillermo Hernández, Javier Prieto and Pablo Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments", Neurocomputing, vol.423, pp.713 - 720, 2021.
[72] Aya Nabil, Mohammed Seyam and Ahmed Abou - Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks", IEEE Access, vol.9, pp.140731 - 140746, 2021.
[73] Fangyao Xu and Shaojie Qu, "Data mining of students' consumption behaviour pattern based on self - attention graph neural network", Applied Sciences, vol.11, no.22, pp.1 - 18, 2021.
[74] Giannakas F, Troussas C, Voyiatzis I and Sgouropoulou C, "A deep learning classification framework for early prediction of team - based academic performance", Applied Soft Computing, vol.106, pp.1 - 17, 2021.
[75] Saeed Ul Hassan, Hajra Waheed, Naif R Aljohani, Mohsen Ali, Sebastián Ventura and Francisco Herrera, "Virtual learning environment to predict withdrawal by leveraging deep learning", International Journal of Intelligent Systems, vol.34, no.8, pp.1935 - 1952, 2019.
[76] Naif Radi Aljohani, Ayman Fayoumi and Saeed - Ul Hassan, "Predicting at - risk students using clickstream data in the virtual learning environment", Sustainability, vol.11, no.24, pp.1 - 12, 2019.
[77] Nida Aslam, Irfan Ullah Khan, Leena H Alamri, Ranim S Almuslim, "An improved early student's performance prediction using deep learning", International Journal of Emerging Technologies in Learning (IJET), vol.16, no.12, pp.108 - 122, 2021.
[78] Bashir Khan Yousafzai, Sher Afzal Khan, Taj Rahman, Inayat Khan, Inam Ullah, Ateeq Ur Rehman, Mohammed Baz, Habib Hamam and Omar Cheikhrouhou, "Student - performulator: Student academic performance using hybrid deep neural network", Sustainability, vol.13, no.17, pp.1 - 21, 2021.
[79] Hajra Waheed, Saeed - Ul Hassan, Naif Radi Aljohani, Julie Hardman and Raheel Nawaz, "Predicting academic performance of students from VLE big data using deep learning models", Computers in Human behavior, vol.104, pp.1 - 34, 2019.
[80] Xizhe Wang, Xiaoyong Mei, Qionghao Huang, Zhongmei Han and Changqin Huang, "Fine - grained learning performance prediction via adaptive sparse self - attention networks", Information Sciences, vol.545, pp.223 - 239, 2021.