How to Cite:
Suresh, B., & Devi, R. R. (2022). Real time multi feature performance analysis model for efficient prediction of student performance using neural networks. *International Journal of Health Sciences*, *6*(S2), 6902–6915. https://doi.org/10.53730/ijhs.v6nS2.6732

Real time multi feature performance analysis model for efficient prediction of student performance using neural networks

B. Suresh
Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Chennai
Corresponding author email: sureshlokshi@gmail.com

R. Renuga Devi
Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Chennai
Email: nicrdevi@gmail.com

***Abstract*---**The problem of student performance analysis and prediction has been well studied. There exist number of approaches to handle this issue which would use variety of features like the involvement in sports, education, academic and others. However, they suffer to achieve higher performance in predicting the performance of the student. To handle this issue, an efficient real time multi feature Performance Analysis Model (MFPAM) is presented in this article. The model consider number of features include the number of sessions the student attended, the number of seminars appeared, number of tests cleared, number of webinars accessed, number of assignments submitted, number of queries produced, number of subjects cleared in each time stamp, number of sports played, number of videos visited, number of extracurricular activities appeared and so on. By analyzing each features in different time stamp of the academic carrier, the method predict the performance of the student. The logs of student have been trained with neural network and at the test phase, the neurons perform analysis on each factor to produce support on different performance factors. At the output layer, the neurons generate set of weight measures towards various class of performance. Based on the weight measures of different class of performance, the method computes the performance weight (PW) based on which the student performance has been predicted. The proposed method improves the performance of student performance prediction with less false ratio.

Manuscript submitted: 27 Feb 2022, Manuscript revised: 18 March 2022, Accepted for publication: 09 April 2022

***Keywords*---**neural network, machine learning, student performance analysis, MFPAM, PW, student data.

Introduction

The recent trend in technology has been used for several scientific applications. Education sector is the one which has higher implications from the development. In recent times, the education society looks forward to get grow with the technology. The students engaged in educational society has several factors and the educational organizations has the responsibility to monitor the growth of the students and should monitor the performance of students. Monitoring the student for their performance is not just about the studies but also depend on various factors. The growth of student should be from various perspectives like extracurricular, sports and other activities. So, the performance analysis of student makes great deflection in their future.

In earlier days, the students were engaged in studies only through the white boards but in now a days, smart classes and online classes has arrived. In this stage, the students studies various subjects through online classes and webinars. Also, they appear in different classes through webinars which makes their performance different. Also, the performance of student is depending on various other parameters. There are number of approaches available in measuring and predicting the student performance. Some of the methods uses the marks and rank scored on various subjects and some other approaches uses various other features like the seminars and other activities. However, the methods suffer to achieve expected performance.

Towards this issue, an efficient real time multi feature approximation model is discussed in this paper. In general, the performance of student should be identified based on the webinars attended. By monitoring so, the involvement of student in listening online classes, can be identified. Similarly, the involvement of student in sports should be considered and the involvement in other social and extracurricular activities must be considered. Further, the performance is depending on how many classes, webinars attended by the students. So the analysis should be performed over the records of the student which requires huge data.

The data mining and machine learning algorithms has great impact in various scientific problems. In this way, the machine learning algorithms can be used in predicting the student performance. The machine learning algorithms are capable of considering missing features and handle higher dimensional problem. When the student data set has higher number of features, then it can be handled by machine learning algorithms. The machine learning algorithms have Genetic algorithm, neural network, fuzzy approaches. Among them, the neural network can be used in our problem to handle the time orient data set. As, the focus of our approach is to analyze the performance of student in different time stamp, the neural network can be used. Because, the student performance in one semester would be differ from others. So, it is necessary to analyze the performance graph of student at different time stamp which would help to find the exact reason and which factor affect the performance of student can be identified. By considering all this, a real time multi feature analysis model for student performance prediction is presented in this article. The detailed approach is presented in this paper.

Related Works

Several approaches are sketched by different researchers towards predicting student performance. This section details some of the methods related to the problem. EKT: Exercise-Aware Knowledge Tracing for Student Performance Prediction [1], design two implementations on the basis of EERNN with different prediction strategies, i.e., EERNNM with Markov property and EERNNA with Attention mechanism. Then, to explicitly track student's knowledge acquisition on multiple knowledge concepts, we extend EERNN to an explainable Exercise-aware Knowledge Tracing (EKT) framework by incorporating the knowledge concept information, where the student's integrated state vector is now extended to a knowledge state matrix.

Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification [2], investigate and demonstrate the effectiveness of our entire approach on seven publicly available and varying datasets. Social Network and Sentiment Analysis: Investigation of Students' Perspectives on Lecture Recording [3], employ social network and sentiment analysis because these methods are useful in examining semi-structured and unstructured social media data. Overall findings suggest students generally view lecture recordings as resources for supplementing live lectures rather than replacing them.

Student Performance Prediction Based on Blended Learning [4], This work is practical and provides enlightenment for learning analysis and individualized teaching in blended learning. Students' behavior in blended learning can be used to predict their learning outcomes, and the implementation method is reproducible. Multiple Features Fusion Attention Mechanism Enhanced Deep Knowledge Tracing for Student Performance Prediction [5], proposes a novel framework for student performance prediction by making full use of both student behavior features and exercise features and combining the attention mechanism with the knowledge tracing model.

Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs [6], propose a framework to automatically analyzing opinions of students expressed in reviews. Specifically, the framework relies on aspect-level sentiment analysis and aims to automatically identify sentiment or opinion polarity expressed towards a given aspect related to the MOOC.

Predicting Student Performance in an Educational Game Using a Hidden Markov Model [7], shows that students' knowledge throughout the intervention can be estimated by time-series analysis using a hidden Markov model (HMM). Toward Understanding Students' Learning Performance in an Object-Oriented Programming Course: The Perspective of Program Quality [8], propose a planned quantitative method for assessing students' gains in terms of programming performance and testing performance. Based on real data collected from students who engaged in our course, we use trend analysis to observe how students' performance has improved over the whole semester. By using correlation analysis, we obtain some interesting findings on how students' programming performance correlates with testing performance, which provides persuasive empirical evidence in integrating software testing practices into an Object-oriented programming curriculum.

Explainable Student Performance Prediction Models: A Systematic Review [9], investigate explainable models of student performance prediction from 2015 to 2020. We analyze and synthesize primary studies, and group them based on nine dimensions. Nonparametric Analysis of the Effect of Knowledge Integration Activities on Third-Year Undergraduate Performance [10], presented a analysis techniques that were used provide novel examples of applications to student data.

An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns [11], propose an unsupervised ensemble clustering framework to use student behavioral data to discover behavioral patterns. The proposed framework extracts behavior features from the two perspectives of statistics and entropy and then combines density-based spatial clustering of applications with noise (DBSCAN) and k-means algorithms to discover behavioral patterns.

Teaching Digital Circuit Design With a 3-D Video Game: The Impact of Using In-Game Tools on Students' Performance [12], investigates the impact of three types of instructional in-game tools on engineering students' performance on digital circuit design tasks. Persistence and Performance in Co-Enrollment Network Embeddings: An Empirical Validation of Tinto's Student Integration Model [13], present a novel methodological approach using graph embedding techniques to capture both structural and neighborhood-based features of the co-enrollment network. In keeping with Tinto's model, we find that not only do these embedded representations of students' social network predict their final grade point average (GPA), but also are able to successfully classify students who dropout.

Academic and Demographic Cluster Analysis of Engineering Student Success [14], A cluster analysis was conducted to understand groups of students based on academic performance and demographic information. Tiered Assignments in Lab Programming Sessions: Exploring Objective Effects on Students' Motivation and Performance [15], proposed approach is a modification of an assignment tiering model found in the academic literature. Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data [16], present a model named Augmented Education (AugmentED) is proposed.

Improving Cross-Modal Image-Text Retrieval With Teacher-Student Learning [17], take full advantage of image-to-text and text-to-image generation models to improve the performance of the cross-modal image-text retrieval model by incorporating the text-grounded and image-grounded generative features into the cross-modal common space with a "Two-Teacher One-Student" learning framework. In addition, a dual regularizer network is designed to distinguish the mismatched image-text pairs from the matched ones.

An Integrated Framework Based on Latent Variational Autoencoder for Providing Early Warning of At-Risk Students [18], proposes an integrated framework (LVAEPre) based on latent variational autoencoder (LVAE) with deep neural network (DNN) to alleviate the imbalanced distribution of educational dataset and further to provide early warning of at-risk students.

Recognition of Students' Mental States in Discussion Based on Multimodal Data and its Application to Educational Support [19], design multimodal analytics to augment the ability to recognize different mental states and found that fusing heart rate and acoustic modalities yields better recognize the states of concentration.

Learning Student Networks via Feature Embedding [20], discuss that the teacher-student paradigm from a new perspective of feature embedding. By introducing the locality preserving loss, the student network is encouraged to generate the low-dimensional features that could inherit intrinsic properties of their corresponding high-dimensional features from the teacher network. The resulting portable network, thus, can naturally maintain the performance as that of the teacher network. Theoretical analysis is provided to justify the lower computation complexity of the proposed method. All the above discussed approaches suffer to achieve expected performance in student performance analysis.

Real Time Multi Feature Analysis Based Performance Prediction Model Using Neural Network:

The proposed multi feature Performance Analysis Model (MFPAM) reads the student data set and applies preprocessing on the data set. The preprocessing scheme eliminates the noisy records from the data set. Further, feature extraction is applied to extract several features like number of sessions the student attended, the number of seminars appeared, number of tests cleared, number of webinars accessed, number of assignments submitted, number of queries produced, number of subjects cleared in each time stamp, number of sports played, number of videos visited, number of extracurricular activities appeared and so on. Extracted features are converted into feature vector and used to perform analysis by training the data into neural network where the network has been designed to take input and produce output at the final layer. It contains number of layers according to the number of time stamp data it has. The output layer neurons returns performance weight on various factors using which the student performance is presented. Based on the weight measures of different class of performance, the method computes the performance weight (PW) based on which the student performance has been predicted. The detailed approach is sketched in this section.

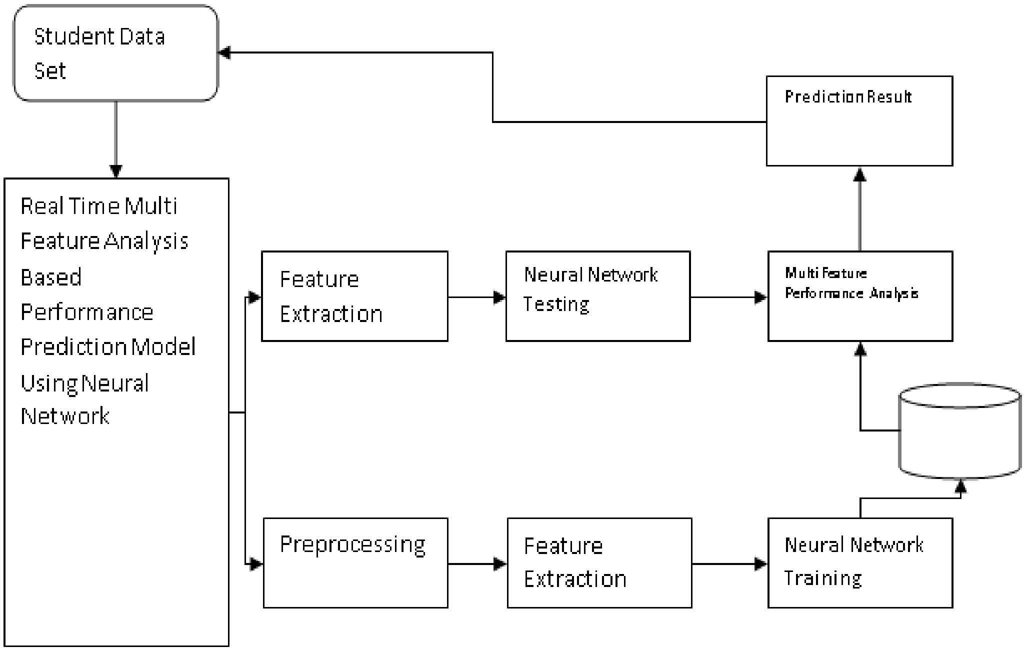

Figure 1: Architecture of Proposed Multi Feature Performance Analysis Model

The functional architecture of proposed multi feature performance analysis model is presented in this Figure 1, where the components of the model are described in detail in this section.

Preprocessing:

The student data set contains number of features and has several entries generated at various time stamp. The method read such data set and finds the features present in the data set. Once all the features present in the data set are identified, then the traces present in the data set are traversed to find the completeness of the data set according to the containment of all the features. If the data tuples identified with incomplete features, then the tuple has been removed from the data set. Such noise removed data has been used towards performance analysis.

Algorithm

Given: Student Data set Stds
Obtain: Preprocessed data set Prds.
Start

Read Stds.
Find features set Fs = $\sum_{i=1}^{size(Stds)}(Stds(i).Features \ni Fs) \cup Fs$
For each trace Ti
If Ti$\in Features(\forall Fs))\ then$
Prds = Prds $\cup\ Ti$

Else
Stds= Stds $\cap Ti$
End
End
Stop

The preprocessing algorithm finds the features of data set and eliminates the set of traces which has missing values and features. Such noise removed data set has been used to perform student performance prediction.

Feature Extraction

The student data contains number of traces generated at different time stamp. To extract the features of the trace, the method first identifies the number of time stamp the records are produced. Once the number of time stamp are produced, then the method split the records under each time stamp. At each time stamp record, the method extracts the features and split according to the time stamp. Generated feature set has been used to perform training.

Algorithm

Given: Preprocessed Data set Prds
Obtain: Time window Feature set Twfs.
Start
Read Prds.
Time window list Twl = $\sum_{i=1}^{size(Prds)}(Prds(i).Timestamp \ni Tw) \cup Twl$
For each time stamp Tw
Initialize feature set Fs = $\sum_{i=1}^{size(Prds)}(Prds(i).Timestamp == Tw)$
For each trace Ti
If Timestamp==Tw then
Feature vector Fv = $\sum Features \in Ti$
Fs = FS $\cup Fv$
End
End
Twl(Tw) = Fs
End
Stop

The feature extraction algorithm extracts various features of data set and split them according to the time window being generated. Such data has been used to perform training towards student performance analysis.

Neural Network Training

The network considered for student performance prediction has been performed in this stage. To perform this, the method finds the size of entire training sample at the initial time stamp. Now the method generates number of neurons at the input layer and initializes each of them with the set of feature extracted. Now, the subsequent layers are generated according to the total time stamp. At last the final layer neuron is generated equal to the input layer. Where the neurons at output layer returns prediction weight on different performance class. Each intermediate layer neurons computes performance score on various features to produce the result to the final layer.

Neural Network Testing

The test sample given has been used in testing the network for predicting the student performance. The method pass the feature being extracted from the test sample to the neural network being trained. The neurons estimates performance support on various activities like Studies, Seminars, number of sessions the student attended, the number of seminars appeared, number of tests cleared, number of webinars accessed, number of assignments submitted, number of queries produced, number of subjects cleared in each time stamp, number of sports played, number of videos visited, number of extracurricular activities appeared and so on. Using them the method computes the value of Session performance Support (SPS), Webinar Access Performance Support (WAPS), Assignment Performance Support (APS), Query Performance Support (QPS), Sports Performance Support (SpPS), Extracurricular Performance Support (EPS). These values are measured on each time stamp according to the feature set belongs to different time stamp traces. Estimated values of all performance support have been returned at the output layer. Obtained values are used to perform student performance prediction.

Algorithm

Given: Test Sample Ts, Neural Network Nn
Obtain: Sps, WAPS, APS, QPS, SpPS, EPS sets.
Start
    Read test sample Ts, and Neural network Nn.
    At each layer l
        Each neuron N
            Estimates Session Performance Support SPS.

$$SPS = \frac{\sum_{i=1}^{size(Fs)} Fv(Time)==l.Time \&\& Fv.SessionAttendance==true}{size(Fs)}$$

            Estimate Webinar Access Performance Support (WAPS).

$$WAPS = \frac{\sum_{i=1}^{size(Fs)} Fv(Time)==l.Time \&\& Fv.WebinarAttendance==true}{\sum_{i=1}^{szize(Fs)} Fs(i).Webinar==True}$$

            Estimate Assignment Performance Support APS.

$$APS = \frac{\sum_{i=1}^{size(Fs)} Fs(i).Assignment==Submitted}{\sum_{i=1}^{size(Fs)} Fs(i).Assignment==True}$$

            Compute query performance support (QPS).

$$QPS = \frac{\sum_{i=1}^{size(Fs)} Fs(i).Queryraised==True}{\text{Number of Sessions Attended}}$$

            Compute Sports Performance Support SpPS.

$$SPPS = \frac{\text{Number of Sports Played}}{\text{Total Sports Conducted}}$$

            Compute Extracurricular performance support EPS.

$$EPS = \frac{\text{Number of Extracurricular event participated}}{\text{Total Activities Conducted}}$$

        End

Compute cumulative value on each and add to set.

$$Spss(l) = \frac{\sum_{i=1}^{size(Fs)} Sps}{Size(Fs)}$$

$$Waps(l) = \frac{\sum_{i=1}^{size(Fs)} Waps}{Size(Fs)}$$

$$Aps(l) = \frac{\sum_{i=1}^{size(Fs)} Aps}{Size(Fs)}$$

$$Qps(l) = \frac{\sum_{i=1}^{size(Fs)} Qps}{Size(Fs)}$$

$$SpPS(l) = \frac{\sum_{i=1}^{size(Fs)} Spps}{Size(Fs)}$$

$$Eps(l) = \frac{\sum_{i=1}^{size(Fs)} Eps}{Size(Fs)}$$

End
Stop

The testing procedure with the neural network towards performance prediction has been presented in the above algorithm which computes various performance support measures and adds to the concern set. Estimated values of various performance supports towards various activities are used to analyze the performance later.

Multi Feature Performance Analysis

In this stage, the method estimates the performance of the student by measuring various support measures. With the performance support values obtained from the testing phase, the method computes the Performance Weight (PW) for various class. According to the performance weight, the method select the class of the student performance. The method classify the student performance under various classes and for each class, the method compute the value of performance weight. Finally, a class has been selected as result.

Algorithm

Given: Sps, WAPS, APS, QPS, SpPS, EPS sets.
Obtain: Class C.
Start
Read Sps, WAPS, APS, QPS, SpPS, EPS sets.
For each class Ci
Compute Performance weight Pw.

$$Pw = \frac{\sum_{i=1}^{size(Sps)} (Sps(i) \&\& Waps(i) \&\& Aps(i) \&\& Qps(i) \&\& Spps(i) \&\& Eps(i)) == Ci.score}{size(sps)} \times 6$$

End
Class C = Choose the class with maximum Pw.
Stop

The above discussed algorithm estimates the performance weight of the user according to various performance scores. Based on that the method finds the class of the student according to the performance.

Results and Discussion

The proposed multi feature performance analysis model has been implemented using advanced java. The method has been analyzed for its performance under various circumstances. The performance of the method has been measured with number of traces and compared with the values of other approaches. The details of evaluation are presented in Table 1.

Table 1: Details of Evaluation

Parameter	Value
Data set	Student Data set
Tool Used	Advanced Java
Number of Features	20
Number of Traces	1 million

The data set used for performance evaluation has been presented in table 1, which has been considered for the performance evaluation under various constraints. The performance of the method has been measured on the following parameters and compared with the results of other approaches.

Table 2: Analysis on performance prediction

Analysis on Performance Prediction %			
	3 lakh	5 lakh	1 million
EKT	69	74	78
HMM	72	77	83
LVAE	75	81	85
MFPAM	84	89	91

The performances of various approaches are measured under different number of records in the data set. The proposed MFPAM model has produced higher performance than other approaches.

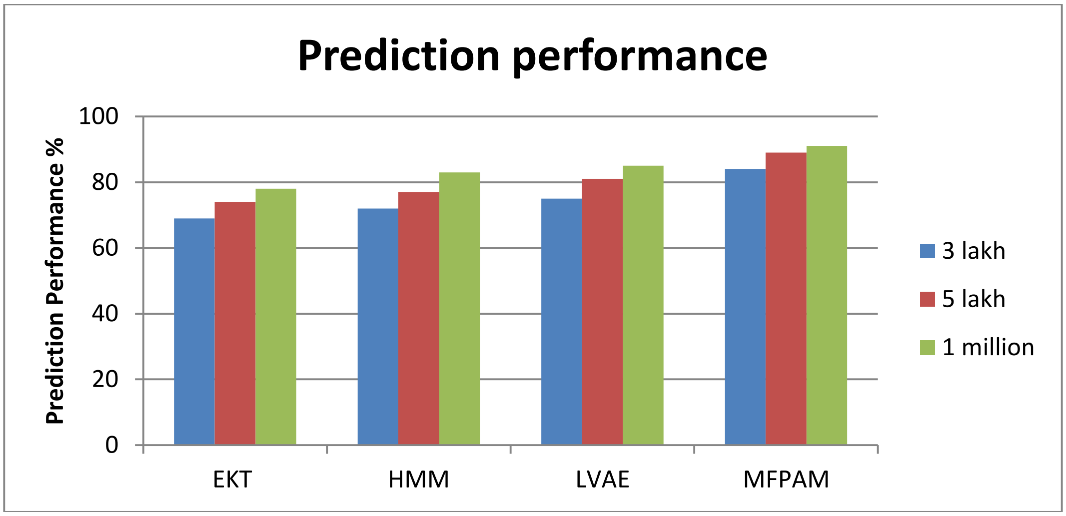

Figure 2: Analysis on student performance prediction

The analysis of student performance prediction has been measured and presented in Figure 2, where the proposed MFPAM model has produced higher performance in student performance prediction.

Table 3: False Ratio on performance prediction

	False Ratio in Performance Prediction %		
	3 lakh	5 lakh	1 million
EKT	31	26	22
HMM	28	23	17
LVAE	25	19	15
MFPAM	16	11	9

The ratio of false classification produced by various approaches are measured and compared in Table 3. The proposed MFPAM model has produced less false ratio compare to other methods.

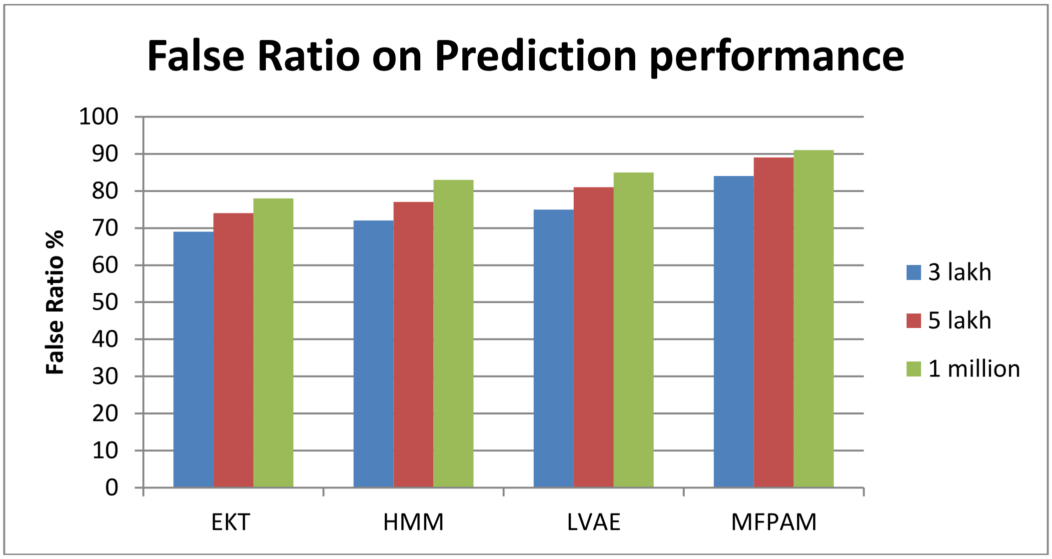

Figure 3: Analysis on False Classification Ratio

The ratio of false classification made by various approaches are measured under various constraints and compared in Figure 3, where the proposed MFPAM algorithm has produced less false ratio than other approaches.

Table 4: Analysis on Time complexity on performance prediction

Time Complexity on Performance Prediction in seconds			
	3 lakh	5 lakh	1 million
EKT	69	74	78
HMM	65	69	73
LVAE	54	61	68
MFPAM	23	39	51

The time complexity introduced by various approaches are measured and compared in Table 4, where the proposed MFPAM model has produced less time complexity in predicting the student performance.

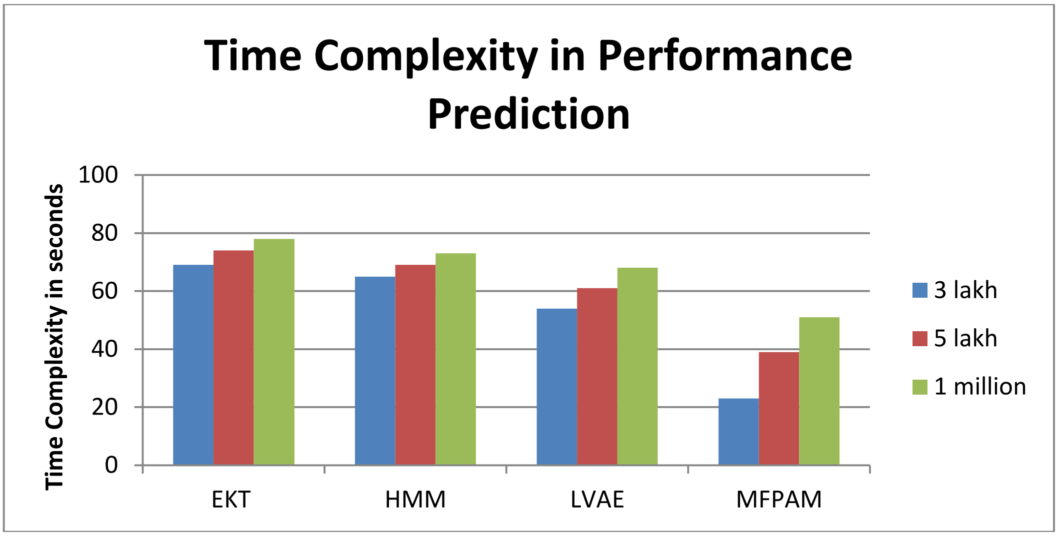

Figure 4: Analysis on Time complexity in Predicting Performance

The performance of time complexity introduced by various approaches in predicting the student performance is measured and compared in Figure 4, where the proposed MFPAM model has produced less time complexity in all the cases.

Conclusion

This paper presented a novel implementation of multi feature performance analysis model with neural network towards predicting the student performance. The preprocessing scheme eliminates the noisy records from the data set. Further, feature extraction is applied to extract several features. Extracted features are converted into feature vector and used to perform analysis by training the data into neural network where the network has been designed to take input and produce output at the final layer. It contains number of layers according to the number of time stamp data it has. The output layer neurons returns performance weight on various factors using which the student performance is presented. Based on the weight measures of different class of performance, the method computes the performance weight (PW) based on which the student performance has been predicted. The proposed method improves the performance of predicting the student performance.

References

1. Q. Liu et al., "EKT: Exercise-Aware Knowledge Tracing for Student Performance Prediction," IEEE (TKDE), Volume 33, Number 1, pp. 100-115, 2021.
2. A. Alshanqiti and A. Namoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification," IEEE, Volume. 8, pp. 203827-203844, 2020.
3. L. M. Nkomo, I. G. Ndukwe and B. K. Daniel, "Social Network and Sentiment Analysis: Investigation of Students' Perspectives on Lecture Recording," IEEE, Volume. 8, pp. 228693-228701, 2020.

4. Z. Xu, H. Yuan and Q. Liu, "Student Performance Prediction Based on Blended Learning," IEEE (TE), Volume 64, Number 1, pp. 66-73, 2021.
5. D. Liu, Y. Zhang, J. Zhang, Q. Li, C. Zhang and Y. Yin, "Multiple Features Fusion Attention Mechanism Enhanced Deep Knowledge Tracing for Student Performance Prediction," IEEE, Volume. 8, pp. 194894-194903, 2020.
6. Z. Kastrati, A. S. Imran and A. Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs," IEEE, Volume. 8, pp. 106799-106810, 2020.
7. M. Tadayon and G. J. Pottie, "Predicting Student Performance in an Educational Game Using a Hidden Markov Model," IEEE (TE), Volume. 63, Number. 4, pp. 299-304, 2020.
8. Q. Sun, J. Wu and K. Liu, "Toward Understanding Students' Learning Performance in an Object-Oriented Programming Course: The Perspective of Program Quality," IEEE, Volume. 8, pp. 37505-37517, 2020.
9. R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," IEEE, Volume. 9, pp. 33132-33143, 2021.
10. C. Robbiano, A. A. Maciejewski and E. K. P. Chong, "Nonparametric Analysis of the Effect of Knowledge Integration Activities on Third-Year Undergraduate Performance," IEEE (TE), Volume. 63, Number. 4, pp. 305-313, 2020.
11. X. Li, Y. Zhang, H. Cheng, F. Zhou and B. Yin, "An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns," IEEE, Volume. 9, pp. 7076-7091, 2021.
12. M. Oren, S. Pedersen and K. L. Butler-Purry, "Teaching Digital Circuit Design With a 3-D Video Game: The Impact of Using In-Game Tools on Students' Performance," IEEE (TE), Volume. 64, Number 1, pp. 24-31, 2021.
13. E. Fincham, "Persistence and Performance in Co-Enrollment Network Embeddings: An Empirical Validation of Tinto's Student Integration Model," IEEE (TLT), volume. 14, Number 1, pp. 106-121, 2021.
14. F. Marbouti, J. Ulas and C. -H. Wang, "Academic and Demographic Cluster Analysis of Engineering Student Success," IEEE (TE), Volume. 64, Number. 3, pp. 261-266, 2021.
15. O. Vojinovic, V. Simic, I. Milentijevic and V. Ciric, "Tiered Assignments in Lab Programming Sessions: Exploring Objective Effects on Students' Motivation and Performance," IEEE (TE), Volume. 63, Number. 3, pp. 164-172, 2020.
16. L. Zhao et al., "Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data," IEEE, Volume. 9, pp. 5453-5465, 2021.
17. J. Liu, M. Yang, C. Li and R. Xu, "Improving Cross-Modal Image-Text Retrieval With Teacher-Student Learning," IEEE (TCSVT), Volume. 31, Number. 8, pp. 3242-3253, 2021.
18. X. Du, J. Yang and J. Hung, "An Integrated Framework Based on Latent Variational Autoencoder for Providing Early Warning of At-Risk Students," IEEE, Volume. 8, pp. 10110-10122, 2020.
19. S. Peng and K. Nagao, "Recognition of Students' Mental States in Discussion Based on Multimodal Data and its Application to Educational Support," IEEE, Volume. 9, pp. 18235-18250, 2021.
20. H. Chen, Y. Wang, C. Xu, C. Xu and D. Tao, "Learning Student Networks via Feature Embedding," IEEE (TNNLS), Volume. 32, Number. 1, pp. 25-35, 2021.