SURVIVAL STUDY ON PREDICTION METHODS FOR EFFICIENT EDUCATIONAL INSTITUTION PERFORMANCE ANALYSIS

C.Vivek[1], P.Tamilselvan*[2]

Abstract

Educational Data Mining (EDM) developed the discipline for discovering the large amount of data from educational settings. EDM collected large amount of data from the educational field. Educational institutions store large amount of data to track students, faculty and courses. The data set comprises personal data and academic student data. To enhance the student lives and faculty as well as universities and organizations using EDM. Several researchers performed in research on educational institution performance prediction. But, the conventional methods were not able to forecast the accuracy and to minimize the amount of time consumed for performing the prediction. To overcome these issues, different educational performance prediction methods are discussed.

Keywords—educational data mining, academic information, universities, organizations, educational institution

I. PREAMBLE

The learning is asignificant role in nation development. The goal of educational institution is to promote the educational procedure through the quality instruction. Every learning organization's success is greatly manipulated by the academic performance of its pupils. During the learning process, the students encounter a number of challenges, including different educational levels, course failure rates, and dropout rates for computer programming classes. The least association rule in the field of education is examined as well asthe educational data sets from e-learning platform are evaluated by EDM using DM processes including categorization, regression, and time series analysis. EDM developed the analytical models to remove the unknown patterns that help in education and learning.

The paper is organized by: In section 2, reviews the existing student performance prediction in educational institution. Section 3 describes the existing educational institution performance analytics. Section 4 provides details on the simulation and the comparative analysis using different metrics. The issues of the existing educational institution performance analytics techniques are covered in Section 5.Section 6 concludes the paper.

OBJECTIVES:

The main aim of my manuscript is to find existing educational institution performance analytics techniques through the crucial metrics as follows.

Prediction Accuracy (PA),
Prediction Time (PT), and
Error Rate (ER).

II. LITERATURE REVIEW

A new machine learning algorithm was designed in [1] for final exam grade prediction with exam grades. But, the proposed algorithm was unable to shorten the time required for feature selection. In [2], the novel hybrid fuzzy approach was implemented the service quality for evaluation. The student expectations in examine value of learning sections were determined by using fuzzy SERVQUAL questionnaires. However, it did not lessen the computing cost.

The probabilistic model was used in [3] to forecast the weighted scores for in-person and online students.

[1,2] Dept of Computer Science, Karpagam Academy of Higher Education, Coimbaotre.
* Corresponding Author

However,it was challenging to reduce the amount of time needed. In [4], an EPDM+ML model was performed in [4] to observe the teacher performance and make recommendation with the consideration of information from student evaluation of teaching (SET).But, it failed to improve the prediction accuracy.

An intelligent system was employed in [5] to improve the user experience in different manner. The student orientation issues were addressed to enhance the performance using Machine Learning (ML) algorithms. Moreover, it failed to diminish the memory usage. In [6], the variability with feature selection techniques were estimated for forecasting course grades by the implementation of bootstrapping model. But, the classification accuracy was not developed by proposed method.

In [7], the predictive model was obtained into forecast the student performance and to analyze the online learning platform features. The machine learning was used to forecast the final grade level of the learner. However, the feature selection accuracy was not enhanced by predictive model. The RBFNN was employed in [8] to prediction of secondary school student performance. The information from the school repository includes the student raw scores and class instructor evaluations. According to the real exam results, the data was classified as pass or fail. The machine learning technique was performed in [9] with the aim of predicting the final student grades in 1 stsemester course through the improvement of predictive accuracy. In order to reduce the results of over fitting and errors in classification, a multi class prediction model was employed. Though over fitting was reduced, the classification time was not minimized.

III. EDM PROCESS

EDMisa method of identifying particular data type from education system with applies the techniques to recognize the students and system. EDM converted the raw data attained from educational systems into useful data to make the data-driven decision. When data has sequential characteristics, the distribution of educational information over time has extraordinary at tributes. AI and ML are used into various sectors like picture classification, ordinary language processing, language acknowledgment, menu script conversion, and field of EDM.

3.1 EDM: ML algorithms based prediction of academic achievement

EDM was coheren tone for classifying the invisible organization within educational data and forecasting student achievements. Machine learning algorithm forecasted final category of apprentice students by practical exam grades. K-nearest neighbor algorithms predicted the student final exam grade. Through the fall semester of 2019–2020, the data set included the academic performance grades of 1854 pupils by Turkish Language-I course in Turkey. Classification accuracy was improved by designed algorithm. The prediction was performed with practical exam rank, branch information and faculty information. The data-driven studies were important one in their learning analysis framework during better teaching and provide to decision-making process. The pre mature for ecasting of students was performed at maximum risk of failure.

3.2 Estimation of educational unit performance based onfuzzy service quality

Educational unites are utilized to resources in effective manner to attain higher service qualities. A novel hybrid fuzzy method was determined by service quality used for educational unit performance appraisal. The assessment and gap examination of student expectation and perception were performed in educational units using fuzzy SERVQUAL questionnaires. Using the fuzzy AHP approach, the size and sub-size weights for SERVQUAL were determined. Through the utilization of fuzzy TOPSIS technique, the educational units were grouped according to the sub-dimensions of service quality. In the fourth stage, service quality ratings were employed in conjunction with the fuzzy DEA approach

to determine the efficacy of educational units.

3.3 Forecasting of Student Performance in Advanced Engineering Mathematics Using Novel Multi variate Copula Models

At an Australian regional university, the probabilistic model was developed for predicting the weighted scores (WS) for advanced engineering mathematics students taking the course in person and online. The semi-parametric D-vine copula used to several questions, tasks as well as marks is validating the predicted WS. The results were obtained and resulted in passing grade in student quiz and assignment scores. For both in-person and online students, the D-vine model and linear regression model provided the grade predictions effectively. The dependent structure for producing arithmetic marks was acquired using combined copula models. The implication supported student achievement and maintenance for verificationshere by engineering necessities with identifying risk. The proposed method guided educator by the effect of engineering problem-solving approximate the student grades.

3.4 Analysis of bootstrap resampling in data pipelines to enumerate the variability in forecasting the performance of pupils

Educators perform an accurate prediction to guess the student maintenance. The bootstrap algorithm was used in data pipelinesto determine the predictive efficiency in student performance predictions. A bootstrapping method was introduced tocompute the variability performance with feature selection methods for predicting course grades. The AUC point evaluates the obtained expect re sampling. DMMs and elastic net regression (GLMNET) express the low variation in AUC. The feature selection methods improved variability in student success prediction by using data pipeline. The bootstrapping was employed to track, observe, and predict the classroom performance. The student presentation in the biology class at the public research university in the Northeastern U. Swas predicted by a DM pipeline.

3.5 Text mining and ML classification based contextual approach for students' assessment examination

An EPDM□+□ML model appraise the teacher recommendations for data from SET. It was utilized to text mining and ML technology with the consideration ofconclusion theory. Itevaluated decisions based on textual data quantification. The pedagogical factors influenced the student recommendations for their teacher in the SET. EPDM□+ML model predicted student recommendation with information about gender, sentiment, and emotional valence. Text mining technique was evacuated thought and sentiment exposed the students within SET. The quantified data performed the covariance analysis and Kruskal Wallis examination to calculate the important features.

3.6 ANNs based students' final performance

In order to forecast the student performance, the EDM is a significant topic. ANNs were determined for efficient student performance prediction within e-learning. The prediction was performed with the student scores. An intelligent system increased the user understanding in diverse manner. The students studied learning in terms of gender, content score and time exhausted in different register courses. Register courses was hard to underst and the input variables in ANNfor variable prediction termed black boxes. An input variable were discussed on output variable prediction.

IV. PERFORMANCE ANALYSIS OF EDUCATIONAL INSTITUTION PERFORMANCE ANALYTICS

During the experiment conduction, the different educational institution performance analytics methods are compared while taking into account the various number of student information as input. For simulation results and student performance data set is taken as an input. Data set predicted to student educational achievement in high school. Experimental evaluation of six techniques, namely machine

learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method, EPDM□+□ML model as well as intelligent systemare implemented using JAVA Platform. The result analysis of existing educational institution performance analytics techniques are carried out by using essential metrics such as PA,PT, and ER.

4.1 Impact of PA

Pais defined as the ratio between the quantity of data points that are predicted accurately and the quantity of data points considered as input.PA is measured in terms of percentage (%) and it is formulated like,

P_Acc=(Number of data points correctly predicted)/(Total number of data points)*100 (1)

By using equation (1), 'P_Acc' is estimated. When PA is maximized other than, the method is said to be further efficient.

Table 1 Tabulation of Prediction Accuracy

Number of Student Data	Prediction Accuracy (%)					
	Machine learning algorithm	New hybrid fuzzy approach	Probabilistic model	Bootstrapping method	EPDM+ML model	Intelligent system
50	82	78	75	71	65	60
100	84	80	77	73	68	62
150	86	83	79	75	70	65
200	89	85	82	77	72	68
250	87	84	80	74	70	66
300	85	82	78	72	67	64
350	83	80	75	70	65	61
400	81	78	72	68	63	59
450	84	81	76	71	66	60
500	87	83	74	73	64	62

The PA for student data ranging from 50 to 500 is shown in the a for ementioned table. While number of student data gets improved, the prediction accuracy gets enhanced or minimized respectively. In order to, measured the number of student data is 250, PA of machine learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method, EPDM□+□ML model as well as intelligent system attained is 87%, 84%, 80%, 74%, 70% and 66% correspondingly.

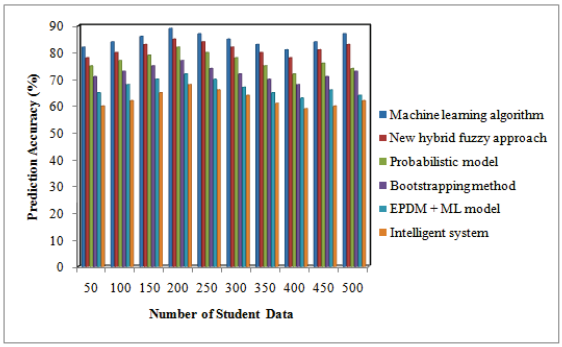

Figure 1 Measurement of Prediction Accuracy

In figure 1,illustrate the PA measure Vs number of student data changing from 50-500. Consequently, the PA of MLalgorithm is comparatively greater than new hybrid fuzzy approach, probabilistic model, bootstrapping method; EPDM□+□ML model plus intelligent system. This is because of forecasting the last test ranks of beginner students. The data-driven studies were employed in learning analysis framework for decision-making process. The early student prediction was carried out at high risk of failure. Consequently, machine learning algorithm increases the PA by 4%, 10%, 17%, 27% and 35% when compared to new hybrid fuzzy approach, probabilistic model, bootstrapping method, EPDM□+□ML model and intelligent system respectively.

4.2 Impact of ER

ER is computed as the fraction of the quantity of data points wrongly predicted and the quantity of data points considered as input. ERis measured in percentage (%). ERis computed as,

E_rate=(Number of data points incorrectly predicted)/(Total number of data points)*100 (2)

From above equation (2), the 'E_rate' is determined. ER minimization indicates the increased efficiency for the approach.

Table 2 Tabulation of Error Rate

Number of Student Data	Error Rate (%)					
	Machine learning algorithm	New hybrid fuzzy approach	Probabilistic model	Bootstrapping method	EPDM + ML model	Intelligent system
50	28	31	35	39	42	24
100	30	34	38	41	45	26
150	27	32	36	38	44	23
200	25	30	33	35	41	21
250	27	33	35	37	43	25
300	30	35	38	40	47	27
350	32	37	40	43	50	29
400	34	39	42	46	52	31
450	36	41	44	49	55	33
500	38	43	48	52	58	35

According to the various quantity of student data, the performance of ER illustrated in tabulation 2.When the number of student data is improved then, ERishigher or lesser respectively. Let us consider, the number of student data is 450, the ERof machine learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method, EPDM□+□ML model with intelligent system achieved by 36%, 41%, 44%, 49%, 55% and 33% correspondingly.

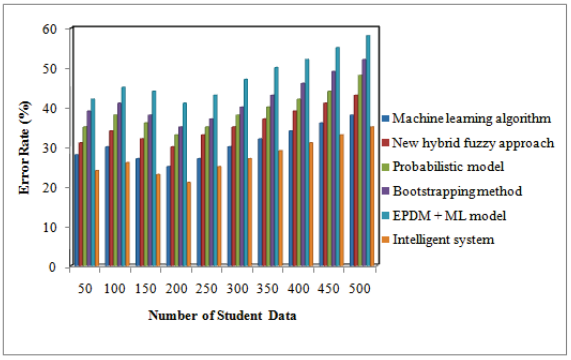

Figure 2 Measurement of Error Rate

In figure 2,shows theERVs number of student data changing by 50-500. Consequently, error rate of intelligent system is compared tolesser than machine learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method and EPDM□+□ML model. This is due to the application of artificial neural networks for predicting performance to establish the student grades. An intelligent system increased their user experience in different manner. The students examining and participating in learning management system was predicted byANNs. As a result, intelligent system reduces the error rate by 11%, 23%, 30%, 35% and 43% when compared to machine learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method and EPDM□+□ML model correspondingly.

4.3 Impact of PT

PT is described as the product of quantity of time required for performing the prediction of one student data and number of student data. It is computed as,

〖Pre〗_T=N*Time consumed to perform prediction of one student data (3)

In above equation (3), the '〖Pre〗_T' is determined. 'N' is a number of student data. When the PT is minimizes, the method is said to be more efficient.

Table 3 Tabulation of Prediction Time

Number of Student Data	Prediction Time (ms)					
	Machine learning algorithm	New hybrid fuzzy approach	Probabilistic model	Bootstrapping method	EPDM + ML model	Intelligent system
50	22	12	18	28	31	36
100	24	14	20	30	33	38
150	26	17	23	33	35	41
200	28	19	25	35	38	44
250	31	22	27	38	40	46
300	33	25	29	40	43	49
350	35	27	31	42	45	51
400	38	29	33	45	48	53
450	40	31	35	47	50	55
500	42	33	37	50	53	58

From table 3, shows the PTVs number of student data varying from 50 to 500. When number of student data isdeveloped, the PTisimprovedrespectively. The number of student data is 350 other thanPTof machine learning algorithm, new hybrid fuzzy approach, probabilistic model, bootstrapping method, EPDM□+□ML model and intelligent systemis achieved35ms, 27ms, 31ms, 42ms, 45ms and 51ms respectively. The graphic depiction of PT is

displayed as the below image.

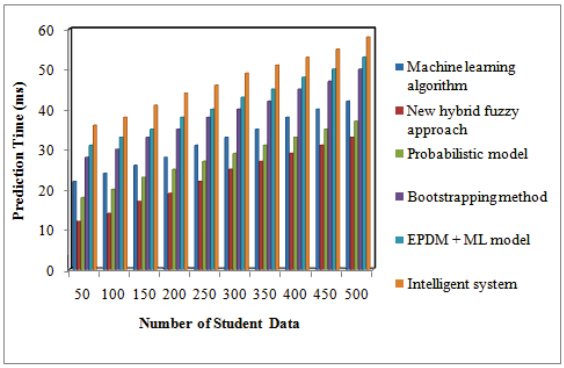

Figure 3 Measurement of Prediction Time

In above figure 3,illustrate the PTVstudent data changing from 50-500. Consequently, the PTof new hybrid fuzzy approach is comparatively lesser than machine learning algorithm, probabilistic model, bootstrapping method,EPDM□+□ML model and intelligent system. This is due to function off uzzy TOPSIS method and fuzzy DEA method to enhance the effective nessby service quality ratings. The designed method improved performance assessment accuracy of educational units with lesser time consumption. As a result, new hybrid fuzzy approach reduces thePTby 30%, 19%, 42%, 46% and 53% when compared to machine learning algorithm, probabilistic model, bootstrapping method and EPDM□+□ML model and intelligent system respectively.

V. DISCUSSION OF LIMITATIONS ON EDUCATIONAL INSTITUTION PERFORMANCE ANALYTICS

A new machine learning algorithm forecasted the undergraduate student grades. The data-driven studies established the learning analysis in higher education for decision-making process. The designed model attained higher prediction accuracy. But, time consumption was not minimized by machine learning algorithm. A new hybrid fuzzy approach performed the educational unit evaluation based on the service quality. A fuzzy DEA method determined the efficiency of educational units. However, it failed to minimizethe computational cost by hybrid fuzzy approach.

Probabilistic model forecasted the weighted scores for face-to-face studies. The designed model guided educator practice through studying combined influences of problem-solving evaluation in student's grades. Other than, there was no reduction in the processing time. EPDM□+□ML examined teacher performance and recommendation from SET. EPDM□+□ML model was determined by combining the text mining and ML strategies with the notion of evocative conclusions. However, the prediction accuracy was not enhanced.

Intelligent scheme is carried the performance predictions based on student scores. Artificial neural network forecasted the student performance based on student scores with higher accuracy. But, the intelligent system was not able to reduce the amount of memory required at an efficient level. Bootstrapping method examined the performance variability as well as preprocessing feature selection techniques for predicting course grades in biology class. However, it failed to improve the accuracy by bootstrapping method.

5.1 Future Work

In the future, ML and Deep Learning (DL) techniques will be utilized for efficiently performing the educational institution performance prediction in a faster manner.

VI. CONCLUSION

The relative study of several educational institution performance predictions is carried out. In study, it failed to improve the prediction accuracy by EPDM+ML. As well, space complexity was not reduced by intelligent system. The hybrid fuzzy approach using computational cost was not minimized. Time consumption was not reduced by machine learning algorithm. The evaluation of existing methods analyzes the outcome of various educational institution performance predictions as well as highlights their problems. As a consequence of the study findings, the ML and DL approaches are employed to enhance the educational institution performance prediction with better efficiency and

minimal time consumption.

REFERENCES

[1] Mustafa Yagci, "Educational data mining: prediction of students' academic performance using machine learning algorithms", Yağcı Smart Learning Environments, Springer, Volume 9, Issue 11, 2022, Pages 1-19

[2] MajidNojavan, AtefehHeidari and DavoodMohammaditabar, "A fuzzy service quality based approach for performance evaluation of educational units", Socio-Economic Planning Sciences, Elsevier, Volume 73, February 2021, Pages 1-10

[3] Thong Nguyen-Huy, Ravinesh C. Deo, Shahjahan Khan, Aruna Devi, AdewuyiAyodeleAdeyinka, Armando A. Apan, and ZaherMundherYaseen, "Student Performance Predictions for Advanced Engineering Mathematics Course With New Multivariate Copula Models", IEEE Access, Volume 10, Pages 45112 - 45136

[4] Kingsley Okoye, Arturo Arrona□Palacios, Claudia Camacho□Zuñiga, Joaquín Alejandro Guerra Achem, Jose Escamilla and Samira Hosseini, "Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification",Education and Information Technologies, Springer, Volume 27, 2022, Pages 3891–3933

[5] TarikAhajjam, Mohammed Moutaib, HaidarAissa, MouradAzrour, YousefFarhaoui, and Mohammed Fattah "Predicting Students' Final Performance using Artificial Neural Networks", Education and Information Technologies, Volume 25, 2020, Pages 1913–1927

[6] Roberto Bertolini, Stephen J. Finch and Ross H. Nehm, "Quantifying variability in predictions of student performance: Examining the impact of bootstrap resampling in data pipelines", Computers and Education: Artificial Intelligence, Elsevier, Volume 3, 2022, Pages 1-12

[7] YudishTeshalBadal and Roopesh Kevin Sungkur, "Predictive modeling and analytics of students' grades using machine learning algorithms", Education and Information Technologies, 2022, Pages 1-31

[8] OlusolaAanuOlabanjo and AshiriboSenaponWusu, "A Machine Learning Prediction of Academic Performance of Secondary School Students using Radial Basis Function Neural Network", Trends in Neuroscience and Education, Elsevier, Volume 29, December 2022, Pages 1-12

[9] SitiDianah Abdul Bujang1, Ali Selamat, Roliana Ibrahim, OndrejKrejcar, Enrique Herrera-Viedma, Hamido Fujita, and Nor Azura Md. Ghani "Multiclass Prediction Model for Student Grade Prediction using Machine Learning", IEEE Access, Volume 9, June 2021, Pages 1-14