Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 1s (2025)

Predictive Analysis System for National Ranking and Accreditation of HEIs

Amrita Sekhon[1*] and Sukhwinder Singh Oberoi[2]
[1]Department of Computer Science, Punjabi University, Patiala, Punjab, India.
[2]Head, Department of Computer Science, Guru Hargobind Sahib Khalsa Girls College, Karhali Sahib, Patiala, Punjab, India.
amritakaursandhu@gmail.com, oberoimca@yahoo.co.in

Article History:
Received: *28-08-2024*
Revised: *14-10-2024*
Accepted: *31-10-2024*

Abstract:
Predictive analytical models have been applied in context of the Indian higher educational institutions dataset. This study is aimed to compare the predictive accuracy level of various machine learning algorithms (multi-class classification prediction algorithms) to assess the grading system applicable to a college based on various parameters outlined by NAAC (National Assessment and Accreditation Council). NAAC is an apex body to assess and accredit HEIs of India. An ensemble classification method is proposed to best predict the academic grading of university-affiliated colleges of India for the NAAC accreditation system. This maiden attempt for using PA (Predictive Analysis) on such institutional credential data would be beneficial to colleges applying for NAAC accreditation to judge their grade prior to final assessment. Thus they would be able to identify their strengths and weaknesses and eventually work towards further improving upon their grade. Such a system would pave the way for designing such predictive systems to assess the institutional rankings for several other assessment systems.

Keywords: academic organization, higher education institution, institutional assessment, machine learning, predictive analysis

1 Introduction

Data mining enables conversion of data to information and information to knowledge. Predictive analysis explores the past and present to gain insight about the future. Predictive analysis holds great significance not only in business and global economy but also in the education sector. Education is directly associated to the socio-economic growth of any country. Higher education institutions (HEIs) like universities and colleges play a crucial role in providing post-secondary progression [1] and for building trained human capital which in turn supports economic development of a country [2]. So quality assessment of such institutions is utmost important to measure the growth and development of a country [1], which in turn is necessary for sustainable global development [3].

Data mining works in the domain of analyzing large repository of data for extracting valuable relations, associations and patterns in it. It acts as a useful tool for transforming data into useful information [4]. It has been found to be useful in health care, banking sector, telecom fraud detection, manufacturing, marketing, social media personalized prompting, surveillance and scientific research, etc. In the field of education it is termed as Educational Data Mining where it is used for course curriculum evaluation, course enrollment prediction, student dropout rate estimation at early stages, student academic performance prediction, fresh graduate employment prediction, alumni salary prediction, learning analytics dashboards, online course satisfaction level estimation of students, etc.

It is very important to assess the quality of higher education providing systems of a nation. It involves qualitative and quantitative assessment of scholastic and non-scholastic aspects of the organization. Performance evaluation is necessary for permanent improvement and quality enhancement of HEIs [5]. There are various agencies that provide national ranking for quality measurement of HEIs in India based on well-defined parameters, criteria and methodologies like NIRF (National Institutional Ranking Framework), QS World University Ranking, Times Higher Education World University Ranking and Outlook-ICARE India Ranking, etc. Accreditation implies providing certification after evaluating the qualitative proficiency of institutions, fora certain period of time. In addition to the above national ranking agencies, the main accreditation agencies for HEIs in India are NBA (National Board of Accreditation), AICTE (All India Council for Technical Education), NAAC (National Assessment and Accreditation Council), etc. NBA and AICTE accredit only professional institutions running courses like hotel management, engineering, pharmacy and computer applications.

NAAC is an autonomous agency of UGC (University Grants Commission) which assesses all types of HEIs, understands their quality status and provides a grade for their accreditation. The method used to assess quality of HEIs by NAAC is similar to that followed by global Quality Assurance agencies. It uses self-assessment method by HEI or System Generated Score (SGS), which is approximately 70% (Quantitative metrics) of total score and peer team visit judgment by NAAC, which is about 30% (Qualitative metrics) of total score to get the final gradation. NAAC awards a grade to an institution based on a seven-point scale which is calculated on the Cumulative Grade Point Average (CGPA) score. HEI having CGPA between 3.51 to 4 gets A++ grade, 3.26 to 3.50 gets A+ grade, 3.01 to 3.25 gets A grade, 2.76 to 3.00 gets B++ grade, 2.51 to 2.75 gets B+ grade, 2.01 to 2.50 gets B grade, 1.51 to 2.00 gets C grade and below 1.5 gets D grade or is not accredited. The NAAC grade is valid for 5 years. After that the HEI needs to get itself re-accredited by going for next cycle of A and A (Assessment and Accreditation). Step one is to provide an SSR (Self Study Report), followed by the procedure of DVV (Data Validation and Verification), which is used to test the SSR. After DVV has been validated and clarified, onsite NAAC peer team visit is conducted for qualitative metrics evaluation.

2 Related Work

This section elaborates the use of predictive analytical models in the education sector to achieve different goals. Since the past two decades, the world has seen an increase in the use of machine learning for educational data mining. Deep learning models have been used to predict student related academic outcomes aiming to improve their learning process [6]. There is not much evidence of the use of machine learning and especially predictive analytics implementation in the field of HEI assessment for accreditation purposes. But many other areas of implementation of ML (Machine Learning) tools can be found in well reputed journals. This review provides ground-work for the present study to bring out unique insights through predictive analysis on higher education institutional data.

Essa and Ayad (Canada, 2012) [7] presented S3 or Student Success System to provide analytical and holistic view of student academic progress. They used learning analytics that is statistical techniques aided with machine learning, to provide a complete solution to identify students who are at risk of failing, design intervention mechanisms for risk mitigation and track efficiency of those interventions using feedback mechanisms. The domains used for learning behavior were participation, attendance, social behavior and course completion. Its key feature was a wide set of data visualizations.

Balakrishnan and Coetzee (USA, 2013) [8] used Hidden Markov Models to identify defining patterns or behavior of Massive Open Online Course (MOOC) students, their tendency to continue studying the course or drop at varied points of time and predict student retention. Individual HMM model was used to predict student behavior, while two Composite HMMs, one using staking and another using cross product digitization were used to predict student retention. Standard binary classification metrics like area under curve (AUC), F1 score, precision, Receiver Operating Characteristic plot (ROC), recall and Matthews correlation coefficient were used to evaluate the predictions.

Jayaprakash et al. (USA, 2014) [9] discussed OAAI, an Open Academic Analytics Initiative project designed to explore the issue of scaling and portability of Learning Analytics technology across different institutions. It was later released under open source license for use by others. It used four classification models for the purpose of training in predictive models namely, decision tree (J48), Naive Bayes, logistic regression and support vector machines.

Jindal and Dutta (India, 2015) [10] presented a use case of Department of Computer Engineering, Delhi Technological University for determining decision making based on predictive analytics. Predictive modeling and data mining was used. Decision trees (C5.0, C4.5, C and RT, J48 and modified C4.5), Weka, artificial neural network and SPSS Clementine were the predictive tools implemented in the project. Decision trees were found to be most suitable to predict the satisfaction level, branch of study and future grades of students.

Ekowo and Palmer (New America, 2017) [11] presented the usage of predictive analysis in the area of higher education in connection with New America's education policy. Targeted student advising (Early alert system and Recommender system), development of adaptive learning courseware, managing enrollment of students, performance based funding, increasing student retention to minimize loss in revenue due to tuition fees and anticipating financial aid are some of the advantages of predictive analytics described with practical examples of US. Challenges of predictive analytics mentioned were labeling, discrimination and stigma based on student demographic data, transparency of student and institutional data, and privacy and security.

Rahman et al. (Malaysia, 2017) [12] proposed a classification model to predict and assess attributes necessary in creating student dataset for suiting selection benchmarks set of employment of fresh graduates. For predicting the employability status, six classification models were selected for the proposed system, viz. Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), Decision Tree (DT), and Neural Networks (NN).

Deuja et al. (Nepal, 2018) [13] used neural networks to predict performance grades of college students. J48 decision tree was implemented for training and testing data models and three-layer perceptron model was implemented in the predictive model to provide an accuracy of 97.12.

Cui et al. (Canada, 2019) [14] reviewed the methodological considerations used in educational data mining by researchers. The merits and demerits of existing applications that were using predictive learning analysis were identified. Recommendations about development, use and evaluation of predictive models were put forth.

Tomasevic et al (Serbia, 2019) [15] analyzed supervised machine learning techniques for predicting student exam performance score. OULAD (Open University Learning Analytics Dataset) was used as dataset comprising of student demographics, engagement and performance data. Three categories of approaches used were similarity based, model based and probabilistic one for performing the analysis using classification (kNN, SVM, ANN, DT, NB and Logistic Regression) and regression techniques(kNN, SVM, ANN, DT, Linear and Bayesian Regression). ANN provided best results for both classification and regression tasks.

Waheed et al. (Pakistan, 2019) [16] used deep artificial neural network on free OULA (Open University Learning Analytics) dataset to analyze student behavior and impact on their performance of their interaction with Virtual Learning Environments. They achieved classification accuracy of 84% to 93% as it outperformed SVM and baseline logistic regression models.

Namoun and Alshanqiti (Saudi Arabia, 2020) [17] performed extensive systematic literature review of intelligent models use to predict student learning outcomes for the period 2010 to 2020. They concluded supervised machine learning models and regression were most used methods to classify student performance, while student online activities, assessment grades and academic emotions of students were the most used predictors of student learning outcomes.

Mengash (Saudi Arabia, 2020) [18] applied ANN, DT, SVM and NB to predict student academic performance before their admission to a university. It used dataset of 2039 students and concluded that ANN performed best with 79.22% accuracy.

Rastrollo-Guerrero et al. (Spain, 2020) [19] reviewed 70 papers to study techniques used for predicting student performance. It showed 70% research papers were on University students and 30% at school level. Supervised learning (SVM being the most used followed by DT, NB and RF) provided accurate and reliable results. Unsupervised learning was not used by researchers due to its low accuracy predicting student behavior. Collaborating filtering algorithms were used for recommender systems than for predicting purposes. They found that neural networks were less used despite providing great precision.

Baig et al. (Malaysia, 2020) [20] reviewed 40 reputed journal publications from 2014 to mid 2019 on big data in education. They found 53% studies on learner performance and behavior, 23% on education system improvement, 15% on educational data warehouse and modeling and 10% on integrating big data on curriculum theme.

Romero and Ventura (Spain, 2020) [21] reviewed the employment of LA (Learning Analytics) and EDM (Educational Data Mining) on educational data. Details about top ten books, journals, conferences, most cites papers and most used LA software, datasets and methods were listed.

Alyahyan and Dustegor (Saudi Arabia, 2020) [22] reviewed literature to provide a collection of guiding principles for educators to implement data mining techniques for predicting academic success of students. Various best practices were put forth along with implementation, and design procedures from related literature were compiled and presented.

Salloum et al. (Sharjah, 2020) [23] reviewed Educational Data Mining trends in educational research. Use of machine learning was used to predict student performance, testing and grading students, enhance retention rate and support teachers by analyzing assessment data. As future trends they predicted use of bigger, adaptable data set, hybridize techniques, improve EDM credibility and compare different methods.

Albreiki et al (UAE, 2021) [24] reviewed literature on educational data mining from 2009 to 2021 based on machine learning techniques for predicting students at risk and dropping out. It was observed that generally small datasets were used and mostly traditional machine learning techniques were used instead of deep learning techniques.

Yağcı (Turkey, 2022) [25] showed comparison of RF, LR, NB, SVM, NN and kNN algorithms to predict student academic performance based on student mid-term exams. They achieved classification accuracy of 74.6% with RF and NN.

Umer et al. (Australia, 2023) [26] surveyed research papers to gain insights about the commonly used data as student performance indicator, commonly used machine learning algorithms for making predictions and evaluation methods for result analysis, their challenges and limitations and future research directions.

Sghir et al. (Morocco, 2023) [27] extensively surveyed literature on PLA in higher education and concluded that deep learning was the choicest method for predictive analysis followed by Random Forest and Gradient Boosting. For evaluating the performance of algorithms confusion matrix was mostly employed followed by Mean Squared Error (MSE), and R-Squared coefficient method.

Table1 summarizes the predictive analytical algorithms used for making different predictions in the field of education.

STUDY	**DATA USED**	**PREDICTION TARGET**	**PREDICTIVE ALOGORITHMS**	**PREDICTION ACCURACY**
Essa and Ayad (2012)	LMS	Student Success	Predictive ensemble methods	Not specified
Balakrishnan and Coetzee (2013)	29,882 MOOC students	Student retention prediction	Composite Hidden Markov Models	AUC score = 0.710
Jayaprakash et al. (2014)	1,739 College student data	Student academic risk prediction	Logistic Regression, Naïve Bayes, SVM/SMO, J48 DT	Recall = 74.5%
Jindal and Dutta (2015)	65,300 AIEEE 2007 dataset	satisfaction level, branch of study	DT, Weka, ANN, SPSS	C5.0=99.95% (enrolment prediction)

	(for enrolment) and Student data of 2 colleges (for future grades prediction)	and future grades of students		C4.5=67% (future grade prediction) C4.5=60% (satisfaction level prediction)
Rahman et al. (2017)	Tracer Study, Ministry of HE data	Employability prediction of fresh graduates	kNN, NB, DT, NN, LR, SVM on Rapid Miner	kNN accuracy=97.78%, Classification error=2.22%, RMS error=0.149+/-
Deuja et al. (2018)	2,016 College students data	Student performance prediction	Multilayer perceptron	Accuracy=97.12%, RMS error=0.0989
Waheed et al. (2019)	32,593 students dataset of OULA	Students at risk of failure prediction	Deep ANN	At risk precision=93%, accuracy=88%, Early Withdrawal precision=96%, accuracy=93%, Distinction precision=81%, accuracy=85%
Tomasevic et al (2019)	3,166 students dataset of OULA	Dropout risk and exam performance prediction	kNN, SVM, ANN, NB, Logistic Regression, DT, BR, Linear Regression	ANN precision=0.9662 ANN RMSE=12.1256
Mengash (2020)	2,039 university (CCIS at PNU) students data	Academic performance prediction in Admission System	DT, SVM, NB, ANN	ANN accuracy=79.22%, precision=81.44% DT recall=80.24%, F1=80.63%
Yağcı (2022)	1,854 University students data	Final exam grade prediction	RF, NN, SVM, Logistic Regression, NB, kNN	RF classification accuracy=74.6%

Table1. Comparative analysis of studies on predictive analysis in higher education sector

3 Problem Statement

In India NAAC accreditation is considered as a benchmark for assessing HEIs educational excellence. It has been made mandatory for HEIs to get it done since 2022 for various financial approvals. Of the seven criteria involved in NAAC assessment, two criteria are related to student services. These two criteria contribute 48 to 50% of the total score. Machine learning has been used in education sector for a long while, but never for predicting the assessment of any institution. This paper tries to figure out the use of various predictive analytical algorithms on institutional assessment and the effect of student services on its overall grading. Thus they would be able to identify their strengths and weaknesses and eventually work towards further improving upon their grade.

4 Methodology

The aim of this study is to find the use and compare the various predictive analytical models to analyze the higher educational institutions for assessing the NAAC (National assessment and accreditation Council) grade. For this purpose data from SSRs (Self Study Reports) of 175 different NAAC accredited colleges of India between the years 2018 to 2023 were used. There are seven different accreditation grades provided by NAAC to various HEIs. The sample size included 25 university affiliated colleges each having grades A++, A+, A, B++, B+, B and C.

The SSR comprises of the Executive summary of the college, which includes the introduction of college, its vision and mission, its SWOC (strength, weakness, opportunity and challenge) record, criteria wise summary, extensive basic profile of the college along with extended profile and the Quality Indicator Framework (QIF). This QIF contains extensive data of five years about the institution which is grouped under seven different criteria, namely Curricular Aspects, Teaching-learning & Evaluation, Research, Innovations & Extension, Infrastructure & Learning Resources, Student Support & Progression, Governance, Leadership & Management and Institutional Values and Best Practices. Each of these criteria further comprise of multiple sub headings called Key Indicators where data under those heads is submitted by each college. It has been mandated by NAAC to upload the SSR of college or university on their respective websites during the accreditation process. Since the scope of this research is confined to student services provided to University affiliated college students, only Criteria 2 and 5 were used for data collection and observation.

Criteria 2 (Teaching-learning and Evaluation) contains data pertaining to student enrolment, students admitted during the last 5 years, number of seats sanctioned for open and reserved categories during those 5 years and percentage and number of seats filled in various reserved categories. The teaching learning process contains student to full time faculty ratio, student centric activities performed, percentage of full time teachers employed against the sanctioned posts and their number, percentage and number of teachers with additional higher degrees apart from basic qualification. The evaluation process includes assessment mechanism and grievance redressal mechanisms used, course and programme outcomes of all courses. Pass percentage, number of final year students appearing and passed and the online student satisfaction survey report about the entire teaching learning process.

Criteria 5 (Student Support and Progression) contains data about students receiving scholarships from various government and non-government sources, skill enhancement initiatives of organization, students benefitted from career counseling and guidance about competitive exams. Student Progression includes data about placement of students and those going for higher education and data about students appeared and qualified competitive exams.

Student Participation/Activities comprises of data about number of awards achieved and number of sports and cultural events organized. The Alumni Engagement part includes the alumni association registration record and its contribution towards the development of the higher education institution.

The weightage of Key Indicators for affiliated colleges for criteria 2 is 350 marks and for criteria 5 it is 140 for UG (Under-Graduate) and 130 for PG (Post-Graduate). Criteria 2 and 5 combined comprise of 490 (UG) or 480 (PG) marks, which is approximately 50% of the total score of 1000 marks. Since the data collected involves two criteria, the aim of this study is to evaluate the score of these two criteria and its effect on the overall NAAC grade of any University affiliated college. The study involves use of predictive modeling techniques. Figure 1 shows the flowchart of methodology used for the research.

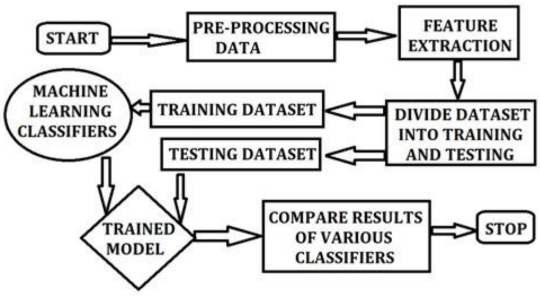

Fig. 1 Workflow of College Classification Model

5 Dataset

The data used as input for this study was collected from different sources like college websites of various NAAC accredited colleges from throughout the country India. Some colleges who did not get good ranking had not posted their SSRs on their websites, so their data was extracted from NAAC data repository containing data regarding grading, score sheets and institutional statistics. The main data specifically comprised of colleges accredited by NAAC during the period from 2018 to 2023. The original dataset consisted of 200 colleges before performing data analysis. Data preprocessing and data cleaning was performed manually as there were a number of attributes that needed to be calculated before entering into the database. In order to have consistent data for proper class balancing, 25 colleges of each grade were selected from throughout the country. Data balancing ensures high classification performance [14]. There are 7 grades for accredited institutions, namely A++, A+, A, B++, B+, B and C. Grade D implies that the institution is not accredited. A total of 175 colleges' data was collected. Each college included institutional data of the last 5 years for assessment, so a total of 875 entities were added to the database. This final dataset of 875 samples used 20 features. Table 2 shows the list of states and number of colleges selected from each state as our sample data.

Number of Colleges for each NAAC Grade								
STATES OF INDIA	**A++**	**A+**	**A**	**B++**	**B+**	**B**	**C**	**TOTAL COLLEGES**
ANDAMAN AND NICOBAR					1			1
ANDHRA PRADESH		1	1	3	2			7
ASSAM		2	2	1	2	2		9
BIHAR				1	1		1	3
CHANDIGARH		1	1					2
CHATTISGARH				1		2		3
DAMAN AND DIU					1		1	2
GUJARAT						1		1
HARYANA	1				1	1		3
HIMACHAL PRADESH		1						1
JAMMU AND KASHMIR						1		1
KARNATAKA	2		2	4	1	1		10
KERALA	6	2	2	1	1			12
MADHYA PRADESH	1		1		6	7	19	34
MAHARASHTRA	2	7	3	8	3	3	3	29
NAGALAND						1		1
NEW DELHI	4	3	2					9
PUNJAB		1	3	3	1	1		9
RAJASTHAN						1		1
TAMIL NADU	6	3	4	1	1	1		16
TELANGANA	1	1						2
TRIPURA						1		1
UTTAR PRADESH	1		1					2
UTTARAKHAND					1			1
WEST BENGAL	1	3	3	2	3	2	1	15
Grand Total	**25**	**25**	**25**	**25**	**25**	**25**	**25**	**175**

Table 2 Dataset of Colleges with corresponding NAAC Grade and States

6 Data preparation

The data obtained from various websites of HEIs contained missing data at times. This missing data was searched at other locations within the same or related files and then noted down manually. Sometimes percentage proportion of a certain field was given and the exact data needed to be calculated. Thus all the pre-processing like data cleaning and data wrangling was done manually before recording the relevant data in the database. The database was split for making predictions in the proportion 80% for training and 20% for testing. This division is necessary for validating the model's ability to generalize on new data [28]. Figure 2 shows the Cluster Map of features depicting the correlation between them.

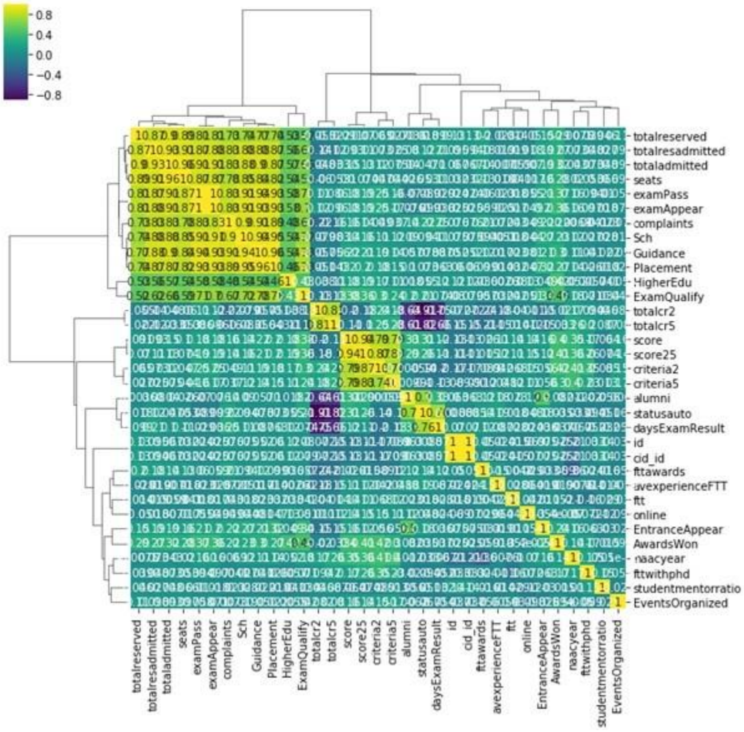

Fig. 2 Cluster map showing Feature correlations

7 Modeling

Classification is the supervised data mining technique used to perform predictive analysis of the NAAC grade of a college. Five traditional classification methods or base models utilized for this research were Decision Tree, Support Vector Machine, Neural Network, Naive Bayes and k-Nearest Neighbors. In addition eight ensemble algorithms like Random Forest, Extra trees, Adaboost, Voting Classifer (DT-SVC), Multinomial Logistic Regression, Linear Discriminant Analysis, Gradient Boosting Classifier and Bagging were used for comparison purposes. Ensemble models were used for making better predictions by combining multiple models than using single model alone. Logistic regression, a famous classification model was not used in this analysis as it works only on binary dependent variables. Bagging ensemble model was proposed to improve the accuracy of prediction. The research was carried out on an open source platform in Python – Django framework (version 4.2.6) for providing data science development environment and data was stored in MySQL file (MySQL Workbench version 8.0.33).

8 Model Evaluation and Results

Thirteen different kinds of Predictive Analytics Classification algorithms were applied on the two criteria chosen from the NAAC Quality Index Framework of Self Study Reports containing data of five previous years of 175 colleges from all across India. Figure 4 shows comparative study of accuracies of the algorithms on four parameters namely Grade based on Criterion 2 alone (C2), Grade based on Criterion 5 alone (C5), Grade based on combined effect of Criteria 2 and 5 (C25) and finally the effect of Criteria 2 and 5 on Overall NAAC Grade of the college. Table 3 shows the accuracies obtained by using various Predictive Analysis tools for various criteria based grading systems.

Sr No	Predictive Analytics Tools	C2	C5	C25	NAAC Grade
1	Decision Tree	25.71%	54.29%	45.71%	51.43%
2	Support Vector Machine	17.14%	22.86%	11.43%	8.57%
3	Neural Network	14.29%	20.00%	28.57%	22.86%
4	Naïve Bayes	31.43%	34.29%	42.86%	42.86%
5	k-Nearest Neighbors	25.71%	34.29%	37.14%	34.29%
6	Random Forest	40.00%	51.43%	51.43%	42.86%
7	AdaBoost	28.57%	37.14%	31.43%	25.71%
8	Voting Classifier	17.14%	22.86%	48.57%	28.57%
9	Multinomial Logistic Regression	22.86%	28.57%	34.29%	28.57%
10	Linear Discriminant Analysis	14.29%	20.00%	34.39%	31.39%
11	Gradient Boosting Classifier	42.86%	45.71%	40.00%	40.00%
12	Extra Trees	**42.86%**	**57.14%**	**54.29%**	51.43%
13	Bagging	31.43%	51.43%	40.00%	**60.00%**

Table 3 Accuracies of Predictive Analysis Tools for various criteria based grading

It was observed that Extra Trees (or Extremely Randomized Trees) algorithm gave best accuracy while predicting the individual grades obtained from Criteria 2 and 5 and combined grade obtained from both the Criteria 2 and 5. But for making the overall NAAC grade prediction, Bagging ensemble method with base estimator as Decision Tree Classifier having 42 random states and 10 n-estimators gave the best accuracy. Support vector machines performed poorly in the research as it is generally used for two group classification problems. It was used only for comparison purposes here. The general conclusion drawn from this research is that ensemble learning methods produce more accurate and robust models by combining individual models. They utilize the strengths and balance out the weaknesses of multiple models and produce more robust and consistent prediction systems. Although 60% is not very high accuracy, but it is fairly good in this case, as we are working here only on two criteria as per the scope of this research and the overall NAAC grade of an organization depends on seven criteria.

9 Conclusion and Future Scope

Higher Educational Institutions have stringent, relevant, productive and consistent performance appraisal for quality assurance and this study would assist them in this direction. The paper presented a comparative analysis of several traditional and ensemble machine learning models for predicting the grade of HEIs of India based on the NAAC assessment and accreditation system. The study proposed Bagging ensemble model as a powerful technique in improving the accuracy of this machine learning model. The model provided an accuracy of 60% of the effect of criteria 2 and 5 on the overall achieved grade of an institution which indicates that although these two criteria contribute about 50% to the grade, but still do not have a big impact on the final grade. The practical implication of this research is that when the HEIs prepare for accreditation or re-accreditation, this system would provide a clear idea of the expected score or grade and the institution would get a chance to make required improvements for upgrading their quality score significantly prior to actual evaluation.

On 29 July 2024, NAAC Reforms 2024 were announced which proposed binary accreditation system and MBGL (Maturity Based Graded Levels). This has been decided so that more HEIs may participate in the accreditation process. The future scope of this research would include using all the seven parameters involved in the NAAC accreditation grade evaluation process to give more accurate assessment prediction. Also implementation of the new MBGL and binary evaluation system would be part of the next phase of this research. The results are expected to improve in the future scope of this research as it would turn into a binary classification problem (accredited or not accredited) instead of the eight grades (A++, A+, A, B++, B+, B, C and D) multi class classification problem.

References

[1] Basheer, N., Ahmed, V., Bahroun, Z., & Anane, C. (2024). Exploring Sustainability Assessment Practices in Higher Education: A Comprehensive Review through Content and Bibliometric Analyses. *Sustainability*, *16*(13), 5799.

[2] Brunner, J. J., Gacel-Avilà, J., Laverde, M., Puukka, J., Rubio, J., Schwartzman, S., & Valiente, Ó. (2012). Higher Education in Regional and City Development Antioquia, Colombia.

[3] Franco, I., Saito, O., Vaughter, P., Whereat, J., Kanie, N., & Takemoto, K. (2019). Higher education for sustainable development: Actioning the global goals in policy, curriculum and practice. *Sustainability Science*, *14*(6), 1621-1642.

[4] Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. In *2016 international conference on data mining and advanced computing (SAPIENCE)* (pp. 91-95). IEEE.

[5] Jalaliyoon, N., & Taherdoost, H. (2012). Performance evaluation of higher education; a necessity. *Procedia-Social and Behavioral Sciences*, *46*, 5682-5686.

[6] Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and information technologies*, *28*(7), 8299-8333.

[7] Essa, A., & Ayad, H. (2012). Improving student success using predictive models and data visualisations. *Research in learning technology*, *20*.

[8] Balakrishnan, G., & Coetzee, D. (2013). Predicting student retention in massive open online courses using hidden markov models. *Electrical Engineering and Computer Sciences University of California at Berkeley*, *53*, 57-58.

[9] Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6-47.

[10] Rajni, J., & Malaya, D. B. (2015). Predictive analytics in a higher education context. *IT Professional*, *17*(4), 24-33.

[11] Ekowo, M., & Palmer, I. (2017). Predictive analytics in higher education. *New America*.

[12] Rahman, N. A. A., Tan, K. L., & Lim, C. K. (2017, October). Predictive analysis and data mining among the employment of fresh graduate students in HEI. In *AIP Conference Proceedings* (Vol. 1891, No. 1). AIP Publishing.

[13] Deuja, R., Karna, R., & Kusatha, R. (2018, October). Data-driven predictive analysis of student performance in college using neural networks. In *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)* (pp. 77-81). IEEE.

[14] Cui, Y., Chen, F., Shiri, A., & Fan, Y. (2019). Predictive analytic models of student success in higher education: A review of methodology. *Information and Learning Sciences*, *120*(3/4), 208-227.

[15] Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, *143*, 103676.

[16] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human behavior*, *104*, 106189.

[17] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*(1), 237.

[18] Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *Ieee Access*, *8*, 55462-55470.

[19] Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, *10*(3), 1042.

[20] Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, *17*, 1-23.

[21] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, *10*(3), e1355.

[22] Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 3.

[23] Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Mining in educational data: review and future directions. In *Proceedings of the international conference on artificial intelligence and computer vision (AICV2020)* (pp. 92-102). Springer International Publishing.

[24] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552.

[25] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.

[26] Umer, R., Susnjak, T., Mathrani, A., & Suriadi, L. (2023). Current stance on predictive analytics in higher education: Opportunities, challenges and future directions. *Interactive Learning Environments*, *31*(6), 3503-3528.

[27] Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and information technologies*, *28*(7), 8299-8333.

[28] Atallah, R., & Al-Mousa, A. (2019, October). Heart disease detection using machine learning majority voting ensemble method. In *2019 2nd international conference on new trends in computing sciences (ictcs)* (pp. 1-6). IEEE.

[29] NAAC (n.d.). NAAC Reforms 2024: Important Announcement –Dt 29th June 2024 Retrieved July 11, 2024, from http://naac.gov.in/index.php/en/