Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Knowledge Based Grade Prediction System using Machine Learning for Higher Education Institutions

Prakash. P[1], Shivanand S. Gornale[1], Rashmi Siddalingappa[2], Satish Kumar[1]

[1]*Department of Computer Science, School of Mathematics and Computing Sciences Rani Channamma University, Belagavi, Karnataka, India*
[2]*Department of Computer Science, Christ University, Bangalore, Karnataka, India*

The enhancement and preservation of standard in the higher education are pivotal for the enduring viability of Higher Education Institutes (HEIs). National Assessment and Accreditation Council (NAAC) in India introduced a new framework for evaluating HEIs in July 2017 based on qualitative and quantitative data analysis and will be assessed and is carried in two ways Data Validation & Verification (DVV) and the onsite peer team visit. The entire Assessment and Accreditation (A&A) process will take the timeline of six to seven months to complete is time consuming and the human intervention. In this proposed work, a predication model using machine learning techniques is developed to assess the performance of HEIs based on the NAAC Criteria within short timeframe and without human intervention. We have used the Multiclass label classification to predict the Key Indicator Qualitative metric score, and the classification based on the total Quantitative & Qualitative Score. The study utilized four distinct algorithms of Machine learning (ML) for classification: Naive Bayes (NB), Random Forest (RT), K-Nearest Neighbors (K-NN) and Support Vector Machine (SVM). The SVM classification technique exhibited the highest accuracy at 97%, followed by Random Forest at 94%, among the four classifiers.
Keywords: Higher Education Institutions (HEIs), Assessment & Accreditation, NAAC, Machine Learning, Support Vector Machine, Random Forest & KNN.

1. INTRODUCTION

In India, the NAAC is the primary regulatory body responsible for maintaining good standards and excellence in higher education. In response to the suggestions of the National

Policy in Education (1986), NAAC was established in 1994. Its main goal is to maintain and improve the quality of higher education while also assessing and accrediting Higher Education Institutions (HEIs). With its main office located in Bangalore, the NAAC actively takes part in the process used by colleges and institutions across the country to monitor and assess academic performance. A combination of internal and external measures is used by the NAAC to guarantee & improve the quality of HEIs [19].

NAAC objectives as listed below:

a. To set up the accreditation and periodic evaluation of HEIs, their units, particular academic initiatives or programs.

b. Encouraging the academic environment in HEIs to promote quality in teaching, research, and learning.

c. Fostering self-evaluation, autonomy, innovations, and accountability in HEIs.

d. Conducting research studies, consulting, and training programs related to quality and

e. Working in conjunction with other higher education stakeholders to assess, promote, andmaintain quality.

The five basic values form the foundations of the NAAC's accrediting structure [20] are mentioned below.

a. Contributing to the development of the country

b. Fostering the global competencies in HEIs students

c. Instilling a value-based system in students

d. Encouraging the use of ICT and

e. Pursuing quality

NAAC evaluation assesses HEI performance according to learning processes and outcome, core curriculum comprehensiveness, teaching & learning methodologies, faculty quality, research activities, infrastructure, availability of learning resources, organizational structure, governance practices, financial sustainability, and provision of learner services [21]. However, these criteria are the primary objectives of a HEIs which is frequently overlooked. Traditionally, using data acquired from the HEIs, assessments were carried for evaluation. The process by which a system can autonomously obtain knowledge from the data and predict outcomes based on this acquired knowledge is referred to as ML, which is a subcategory of artificial intelligence. Various ML techniques can be employed using past data to predict the A&A outcomes of HEIs. These machine learning models can detect potential failures in HEIs and intervene proactively to prevent them.

These studies collectively highlight Machine Learning (ML) and Artificial Intelligence (AI) transformative potential in education, enabling educators to enhance learning experiences, support student success, and improve outcomes through timely and tailored interventions. The objective of this research work is to advance objectivity and efficiency in institutional assessments by utilizing machine learning to revolutionize the NAAC grading procedure. The study looks for major indicators of NAAC grades by evaluating a large amount of educational data, offering useful information for ongoing quality improvement.

The key proposed works contributions are.

- A novel algorithm to predicate the Qualitative metrics (QlM) score based on the given Quantitative metrics (QnM) score and finally with the grade predication.
- The HEI grade is obtained within the short time compared to existing process.

The objective of this research is to create reliable predictive models for evaluating higher education institutions by predicting the qualitative score based on the quantitative score and HEI grade using artificial intelligence techniques. These models aim to provide valuable insights to stakeholders for effective resource planning and informed decision-making by analysing diverse factors such as academic performance indicators, faculty credentials, student demographics, financial allocations, and accreditation standing. Through this endeavour, the objective is to contribute to the advancement of predictive analytics in higher education, fostering excellence and continuous progress in academic institutions.

The rest of paper is followed by study of existing research in Section II, the proposed methodology and experimental findings in Sections III and IV, conclusions and future scope in Section V.

2. RELATED WORK

Numerous studies are currently being conducted on ML applications in the education, exploring a range of approaches. These innovations have the potential to improve educational results and learning environments.

Pinto, Agostinho Sousa, et al. [1] give an in-depth, methodical analysis of the ways in which machine learning (ML) is changing tertiary education, emphasizing its effects on predictive analytics, individualized learning, and administrative efficiency. A variety of machine learning applications are examined in this paper, including predictive models for student performance and adaptive learning systems. It ends by going into the possible advantages and difficulties of using ML in educational settings, including ethical questions and data protection issues.

Alyahyan and Duştegor [2] have examined the main variables and predictive models that are utilized to forecast academic achievement in postsecondary education, with an emphasis on student data. Using machine learning algorithms and ongoing data monitoring to enhance predictions are two of the best practices that are highlighted in this study. The significance of early intervention and tailored support in improving student outcomes is underscored by the authors.

Cervera, Salcedo Parra, et al. [3] have provided a forecasting model that uses machine learning to estimate how well college students will perform on the ICFES exam. The study investigates how various machine learning algorithms can raise the precision of academic forecasts, assisting in the identification of pupils who are at-risk and enhancing educational results. Their model illustrates the possibility of data-driven decision-making for resource allocation and learning process optimization.

Singh and Pal [4] have examined how to increase the precision of student performance prediction by utilizing ensemble methods and machine learning algorithms. Their study shows that mixing different algorithms—like support vector machines and decision trees can produce predictions that are more accurate. In identifying kids who are at danger and directing focused educational interventions, the authors emphasize how beneficial these cutting-edge tools are.

Villegas-Ch, Roman-Cañizares, et al. [5] has suggested enhancing the concept of online education by incorporating data analysis and machine learning into a learning management system (LMS). The study illustrates how these technologies may enhance course delivery optimization, forecast student performance, and personalize learning experiences. The findings demonstrate how combining data analysis and ML can improve education's overall efficacy and enable the creation of adaptable learning pathways.

Latif et al. [6] have examined how machine learning might be used in higher education to evaluate student achievement through the analysis of online activity logs. This study investigates the predictive potential of digital interactions between students and Learning Management Systems (LMS) for academic achievements. According to the research, machine learning models have the potential to improve student assistance and success rates in online learning settings by offering precise assessments and timely interventions.

Hussain et al. [7] have examined how machine learning can be used to forecast student challenges using information gathered from class sessions. The study evaluates several algorithms to discover trends in student behavior that indicate issues in comprehension and engagement. Results show that ML models are able to predict possible learning barriers with high accuracy, which allows teachers to improve student support tactics and intervene in the real time.

Zhai et al. [8] have examined the application of AI in education from 2010 to 2020 and assess how it affects administrative, teaching, and learning procedures. By classifying AI applications into intelligent tutoring systems, adaptive learning environments, and administrative tools, the study highlights improvements in efficiency and personalized learning. Challenges including data protection, ethical issues, and the requirement for successful AI technology integration in educational contexts are also covered in the review.

M. B. Musthafa et al. [9] have introduced a predictive model for assessing university accreditation. The dataset utilized in this study was sourced from the Higher Education Database (HEDB), a repository maintained by Center of Data and Information at Ministry of Research and Technology & Higher Education in Indonesia. The methodologies investigated in this research encompass the KNN and NB algorithms. The KNN algorithm demonstrated a classification accuracy of 95.2% when utilizing a k value of 1, while the Naïve Bayes algorithm achieved an overall accuracy of 70% through 5-fold cross-validation.

Fatek Saeed and Professor Anurag Dixit [10] have introduced a Machine Learning-based Decision Support System approach aimed at aiding the Assessment and Accreditation (A & A) council through the utilization of probability methods and classification using the Naive Bayes algorithm. The authors developed a unique framework known as YAC-Dss, with "Y" representing Yemeni, "A" representing Accreditation, "C" representing Council, and "Dss" representing Decision Support System. The primary objective of this framework is to categorize institutions based on their adherence to national Higher Educational standards in Yemen. Saeed and Dixit curated a proprietary dataset comprising information from various public and private institutions, which was categorized into three classes: Yes, No, and Rep. Subsequently, they computed eleven measures of quantity and quality, and trained the data using the Naive Bayes classification technique. The classification process assigned probability values to each class, resulting in accuracies of 1.9% for the Yes class, 42.3% for the No class, and 55.8% for the Rep class.

Yuri Vanessa Nieto et. al. [11] have emphasized application of ML in decision making within HEIs. They employed supervised classifiers to forecast grades, utilizing a dataset of 6100 students from a public university encompassing five distinct engineering programs. The dataset was used to train three ML models: RF, Logistic Regression and Decision Tree. Evaluation of algorithm performance indicated that RF on top with an accuracyrate at 84.11%, closely followed by LR at 84.02%, and DT with the lowest accuracy rate of 83.92%.

By observing the work carried out in the field of Assessment and Accreditation (A&A) there is a need of more automation in the process. In view of this we have proposed a model which will predict the grading system by looking into the quantitative and qualitative metrics.

Proposed Methodology

The entire approach of the suggested study for the performance prediction of HEIs based on the NAAC and A&A. The generalized proposed methodology architecture is shown in Fig. 1.

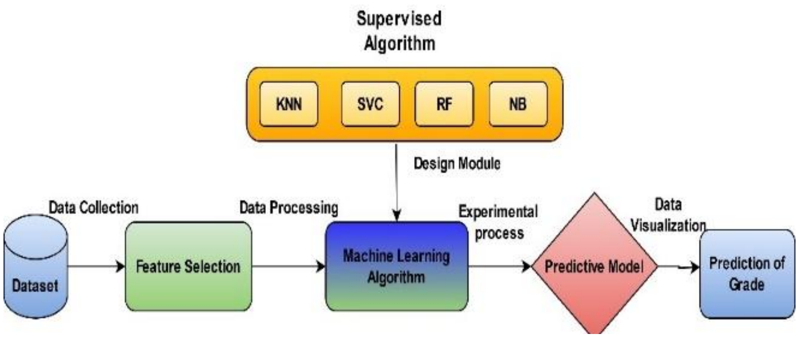

Fig. 1. The architectural diagram of Proposed Methodology

NAAC CGPA & Grade

The database which consists of 1200 HEIs. Every institution is evaluated using a four-point scale (0-4) comprising of QlM & QnM. The Institutional Grade Sheet which is composed of QlM (30%) and QnM (70%). Finally, the CGPA will award to the institutions.

The computation of the institutional CGPA comprises the

a) Key Aspect wise Weighted Grade Point (KAWGP)

It is computed by multiplying a Key Aspect's preset Weightage (W) by the corresponding Key Aspect-wise Grade Points (KAGP) [12]. After physical verification, the peer team assigns the KAGP value, which is a whole integer between 0 and 4. The formula utilized for this is

$$KAWGP_i = KAGP_i \times W_i \qquad (1)$$

b) Criterion wise Weighted Grade Point (C_rWGP),

CrWGP can be compute by two different ways

i. By dividing a criterion's total KAWGP by the total Key Aspect Weightages of that criterion.

$$\mathrm{C_rWGP_i} = \frac{\sum_{i=1}^{n}(\mathbf{KAWGP})_{\mathbf{i}}}{\sum_{i=1}^{n}\mathbf{W_i}} \qquad (2)$$

ii. The method involves dividing the Weighted Grade Point (CrWGP) according to a given criterion by the total weight assigned to it.

$$\mathrm{C_rWGP_i} = \frac{(\mathrm{KAWGP})_{\mathrm{i}}}{\mathrm{W_i}} \qquad (3)$$

c) Criterion wise Grade Point Average (C_rGPA).

The CGPA of the institution is calculated by dividing total of seven CrWGP by the total of the predetermined weights for each of the seven criteria. The institutional CGPA will be used to determine the institution's grade and accrediting status. A, A+, A++, B, B+, B++, C, or D are the letter grades that are determined based on the CGPA, which is as well taken into consideration as the institution's grade.

$$\text{Final CGPA} = \frac{\sum_{j=1}^{7}(\mathrm{C_rWGP})_{\mathrm{j}}}{\sum_{j=1}^{n}\mathrm{W_j}} \qquad (4)$$

Where

i -Key aspect
j - Criteria
n - Number of the Key aspects in criterion.

$\sum_{i=1}^{n}(\mathrm{KAWGP})_{\mathrm{i}}$ - Total of the assigned KAWGP of criterion
$\sum_{i=1}^{n}\mathrm{W_i}$ - Total of the pre-established Weights of that Criterion's Key Aspect

The Final Grade
The ultimate grade is given on a 7-point grading system, the same as indicated in Table I, based on CGPA that the institution was able to acquire with a maximum score of 4.00. The term "seven point" refers to seven letter grades that are corresponding to 7 distinct scoring ranges [13].

TABLE I. HEIS CGPA WITH GRADE AND STATUS

Range of Institutional Cumulative Grade	Letter Grade	Status
3.51-4.00	A++	Accredited
3.26-3.50	A+	Accredited
3.01-3.25	A	Accredited
2.76-3.00	B++	Accredited
2.51-2.75	B+	Accredited
2.01-2.50	B	Accredited
1.51-2.00	C	Accredited
≤ 1.50	D	Not Accredited

HEIs receiving a letter grade of "D" are conceptually classified as having a CGPA of 1.50 or lower. The NAAC will also notify and insinuate these unqualified institutions as “Assessed and Found not qualified for Accreditation."

Data Description

The HEIs mark dataset was gathered from Bengaluru's NAAC. Between January and September of 2023. 1200 different HEIs marks for grades ranging from A++ to C were included in the performance data. The dataset's lone predictive variable is HEIs marks. Using the NAAC Grading System, the possible values of the HEIs grade are A++, A+, A, B++, B+, B, and C [22]. The maximum result that a HEI can earn in this case is A++, and the lowest is C. 1200 data items representing HEI marks in various Key Indicators during the A&A process made up the dataset. There were exactly 274 data items from B grade in the dataset, which made up most of the data. The dataset contains quantitative data (22 Metrics) will be numeric value for which then benchmarks will be applicable, according to assigned value from 0 to 4 and qualitative data (34 Metrics) will be descriptive for which marks will be assigned value from 0 to 4.

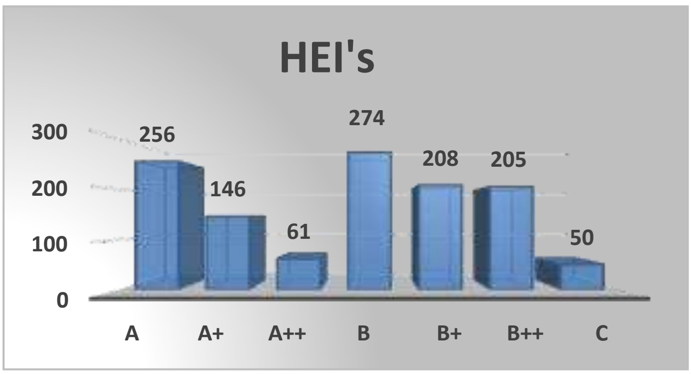

Fig. 2. Frequency of each Grade class in dataset.

Data Processing

In data preprocessing, redundant entries are eliminated to avoid repetition, incomplete data are deleted, and cells are formatted consistently for readability and simpler analysis. The dataset was preprocessed before being fitted into the models to guarantee the best possible performance from them [14]. Standard scalar for features and label encoder for labels were employed in grade classification. The purpose of the Label Encoder was to convert numerical values such as 0, 1, 2, 3, 4, 5 & 6 with A, A+, A++, B, B+, B++, and C respectively.

Feature Selection

We have selected the features i.e. QnM and QlM, because the existing methodology grade depends on the total score of all seven criteria QnM and QlM. For each Key indicator there will be QnM & QlM.

The NAAC 7 criteria i.e. curricular aspects, teaching, learning, and evaluation, research, innovation, and extensions, infrastructure and learning resources, student support and progression, governance, leadership, and management and institutional values & most excellent practice may be important components [15], which includes the 32 Key Indicators. It is easier to evaluate and forecast institutional performance when pertinent aspects are the focus.

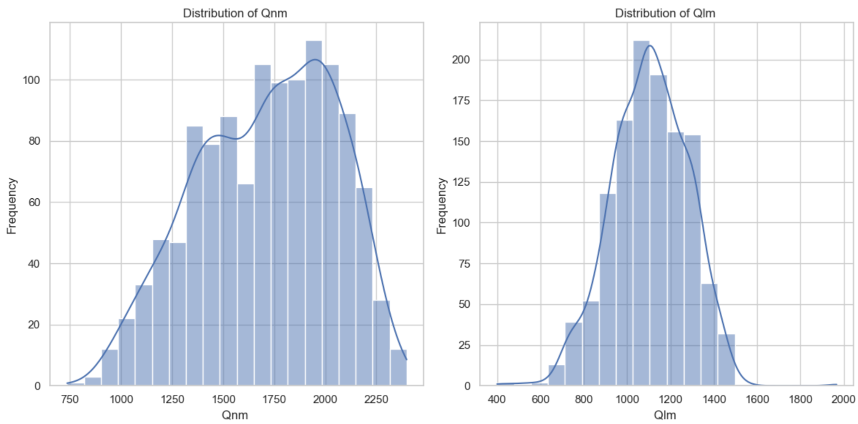

Fig. 3. Frequency of Distribution of QnM & QlM metrices

Dataset Split

To achieve robust model evaluation in our study, we have taken the 1200 data set and we always want to split into a 20:80 ratio. 20% for testing and 80% for training. By studying patterns and correlations within the data, the ML model is constructed and optimized with the aid of training set. We can analyze the model's performance on untested data using the testing set, which is kept apart, giving us an objective evaluation of its accuracy and capacity for generalization.

Hyper parameter Tuning

To overcome model over fitting, hyper parameter tuning was implemented using the Grid Search method, enabling the optimization of relevant parameters for enhanced analysis. The study employed four supervised classification techniques, namely SVM, K-NN, NB, and RF, with Grid Search applied to each model.

Classifier	Parameters Explored	Optimal Parameters	Notes
SVM	'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], kernels: ['rbf', 'linear']	{'C': 100, 'gamma': 0.1, 'kernel': 'rbf'}	Competitive performance achieved
K-NN	'n_neighbors': [3, 5, 7, 10], 'p' (distance metric): [1 (Manhattan), 2 (Euclidean)]	{'n_neighbors': 5, 'p': 2}	Optimal configuration for best performance
NB	N/A (GaussianNB method)	N/A	Utilized GaussianNB for dataset analysis
RT	'n_estimators' : [100, 200, 300], 'max_features': ['auto', 'sqrt', 'log2'], 'max_depth': [4, 6, 8, 10, 12], 'criterion': ['gini', 'entropy']	{'criterion': 'entropy', 'max_depth':12, 'max_feature s': 'sqrt', 'n_estimators ': 200}	Resulted in optimal configuration

The performance and resilience of the classification models were improved by this methodical hyperparameter tuning strategy, confirming its effectiveness in enhancing model training and enhancing predicting abilities in the research setting.

In the proposed study, four conventional ML algorithms were utilized for the analysis of dataset. Firstly, the probabilistic NB algorithm is used to train the data with Gaussian NB parameters

- Precision: It is a measure that provides the ratio of actual positives to the total number of positives the model predicts.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{5}$$

Where:

TP = True Positives (correctly predicted positive cases)

FP = False Positives (incorrectly predicted positive cases)

- Recall: It is a measure that the percentage of actual positive examples in the collection that are true positive predictions.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

Where:

TP = True Positives (correctly predicted positive cases)

FN = False Negatives (actual positive cases that were incorrectly predicted as negative)

F1 score: It is the precision and recall harmonic mean. It balances the trade-offbetween precision and recall by providing a single metric that combines both.

$$\text{F1 Score} = 2\ X\ \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \tag{7}$$

When recall and precision are both very good, the F1 score is maximum.

Receiver Operating Characteristic (ROC) Curve

When assessing the effectiveness of various classifiers, the ROC & their Area Under the Curve (AUC) values are crucial. Ability to Distinguish how well a classifier differentiates between the positive and negative classes is shown by the ROC curve. A higher AUC value denotes a classifier that performs better overall.

Using ROC plots, we may determine the following:

- The trade-off between False Positive Rate and True Positive Rate (Recall) at the various thresholds.
- The AUC, which measures the model's largely discriminative capacity.

A curve nearer the top-left indicates greater performance, this indicates how well the model performs in comparison to random guessing.

Classification Models

We used four distinct machine learning classification approaches in this work: NB, RT, K-NN, and SVM.

1) **K-Nearest Neighbors**: In K-NN, sample data point class is defined by the number of nearest neighbors to be taken into consideration, which is decided by the nearest neighbor using the k-value [16]. In the suggested study, the Minkowski distance formula was utilized to determine the separation between an observation and the centroid.

Based on the features of nearby data points, KNN provides a versatile and comprehensible method for forecasting in data analysis.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (8)$$

Where, p are the given point coordinates with a neighbour q.

2) **Support Vector Machine**: SVM organizes data points into different classes by creating a hyperplane, or decision border, in an N-dimensional space that maximizes margins between classes [17]. The equation for predicting a new input (X) for the linear kernel of SVM is as follows:

$$f(X) = Sum(X * Xi) \qquad (9)$$

Here, Xi is the support vector that created the hyperplane using training data, and X represents the individual subject's marks as a feature vector.

3) **Random Forest**: RT also provides insightful information on the significance of features, which helps with the interpretation and improvement of models. When it comes to data analysis prediction, RT is a dependable and strong method that may produce precise forecasts in a variety of fields [18]. They work by recursively separating the data into subgroups depending on the most informative attributes, establishing a tree-like model of

decisions. Two popular dividing criteria are Entropy and Gini Impurity.
GINI Impurity: Used for Classification

$$Gini(D) = 1 - \sum_{i=1}^{c} p_i^2 \qquad (10)$$

Entropy:

$$Entropy(D) = -\sum_{i=1}^{c} p_i \, log_2(p_i) \qquad (11)$$

Where, pi is the proportion of samples belonging to class i in the dataset D.

4) **Naive Bayes**: Bayes theorem determines the likelihood that an event will occur based on the likelihood that another event has already happened. In mathematical terms, it is expressed as the following equation.

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} \qquad (12)$$

Where, (E|F) is the class E posterior probability with characteristics F.

(F|E) is the probability that characteristics F will exist given class E.

(E) is class E's prior probability.

(F) It features F's prior probability.

3. EVALUATION METRICS RESULTS

The effectiveness of each classifier was evaluated using four evaluation matrices: recall, precision, F1 Score & accuracy. The performance was evaluated by testing those classification models on 240 testing data items. In test dataset, there were 49 data items from the A category, 34 from A+, 07 from A++, 47 from B, 37 from B+, 53 from B++, and 13 from the C category. F1-score for each classifier in each performance category are shown in Table II. Here, SVM achieved the highest f1-score 0.99 in predicting category 4(B+), followed by 0.98 in predicting category 3(B) and followed by 0.97 in predicting category 0(A). Average highest f1-score was attained in category 2(A++) and the lowest average was in category 6(C).

Table II. F1 SCORES FOR EACH RESULT CLASS PREDICTION

Class Name	F1 Score			
	SVM	Random Forest	KNN	Naive Bayes
0 (A)	0.97	0.96	0.95	0.83
1 (A+)	0.96	0.95	0.96	0.63
2 (A++)	0.93	0.93	0.93	0.77
3 (B)	0.98	0.93	0.92	0.76
4 (B+)	0.99	0.91	0.86	0.60
5 (B++)	0.92	0.95	0.90	0.74
6 (C)	0.92	0.87	0.82	0.56

In four classifiers, the SVM performed highest from others with an accurateness of 97% and an average weighted F1 score of 0.96, RT was in the second with an accurateness of 94% and an F1 score of 0.93, Naive Bayes performed the lowest accurateness of 73% and an F1 score of 0.71. The value of four valuation matrices for each classifier. Recall, Precision, F1-score and cross validation accuracy for each classifier in each performance category are shown in the Table III.

TABLE III. PERFORMANCE EVALUATION FOR EACH CLASSIFIER

Classifier	Precision	Recall	F1-Score	Cross Validation Accuracy
SVM	0.96	0.96	0.96	97%
RT	0.94	0.93	0.93	94%
K-NN	0.92	0.90	0.90	91%
NB	0.79	0.68	0.71	73%

A confusion matrix is a way of showing how many of the model's predictions were true and inaccurate, giving a clear visual representation of model's performance. Finding opportunities for improvement in decision-making and classification accuracy is made easier by analyzing these indicators. Here Class 0, 1, 2, 3, 4, 5, 6 represent the NAAC grades A, A+, A++, B, B+, B++ & C respectively.

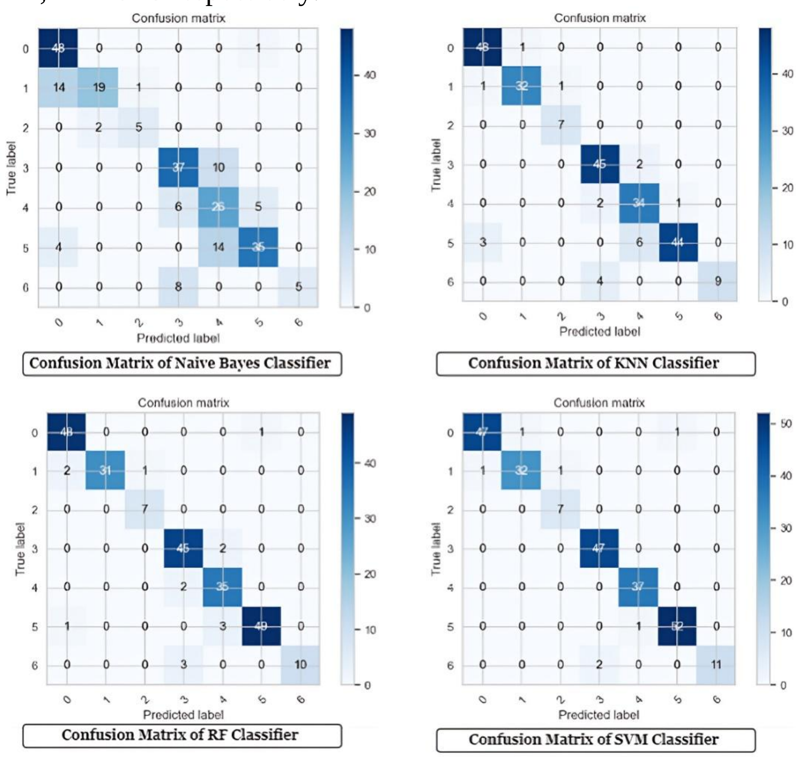

Fig. 4. Confusion Matrix

We can determine which model performs best by comparing the ROC curves of several classifiers, such as NB, K-NN, RT & SVM in Fig. 5. With the greatest AUC value of 1 SVM followed by 0.99 in this instance, RF and K-NN are the most successful in differentiating across classes.

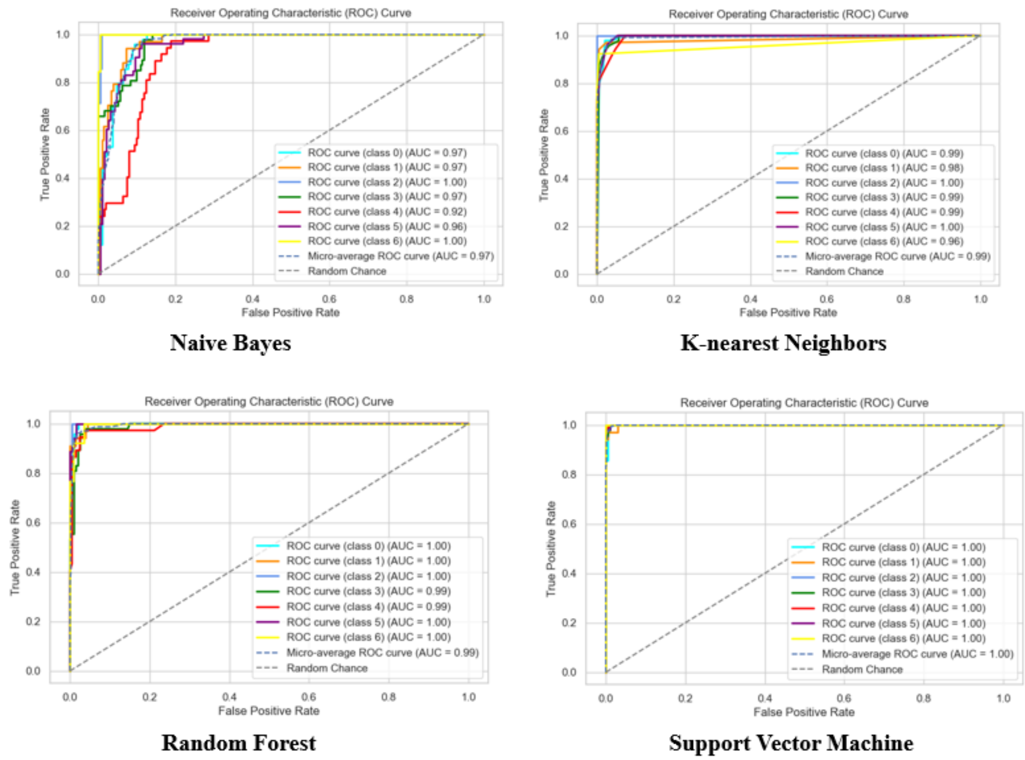

Fig. 5. ROC of each Classifiers

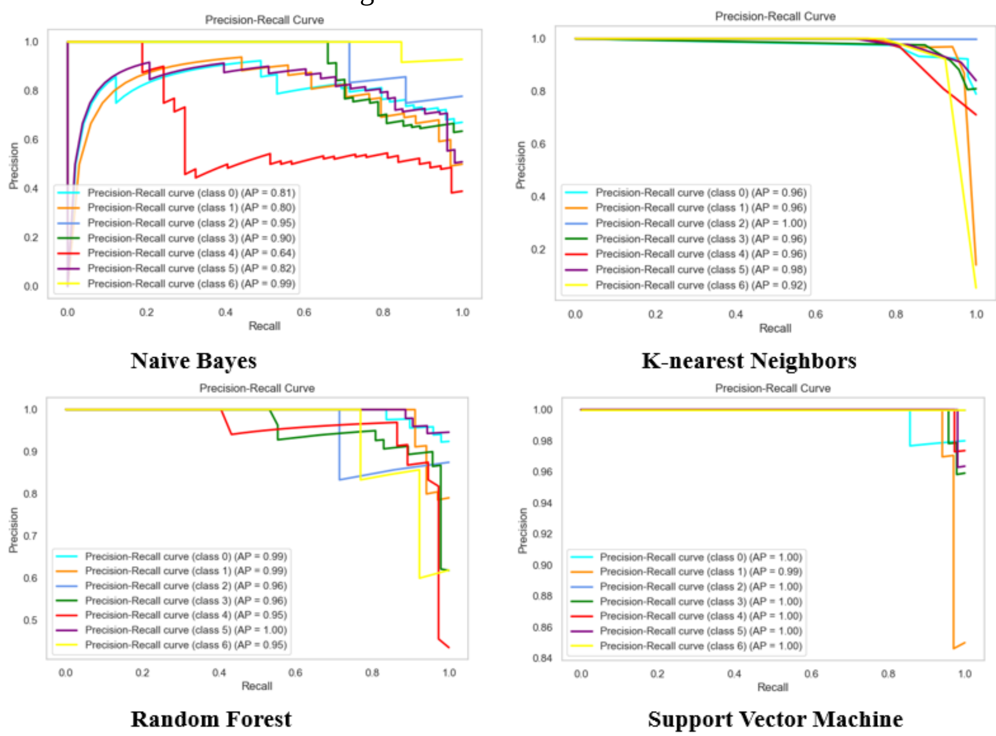

Fig. 6. Precision vs Recall Curve of each Classifier

The Precision-Recall curve illustrates how threshold settings affect recall and precision. While a model with high recall and poor precision detects most positive cases but also includes a large number of false positives, a model with the high precision and low recall is accurate when it predicts positive cases but misses many real positives. With the greatest Precision-Recall curve value of 1, SVM followed by 1 in this instance almost for all the classes in the Fig. 6.

4. CONCLUSION AND FUTURE WORK

The current system requires the significant quantity of time and human participation to declare final grade, which could lead to bias during the onsite visit. With precision and impartiality, the suggested model can predict grades in less time and without the need for human participation. The proposed model is being able to raise the grade in the early stage and with a minimum time. The suggested models showed good performance with a very short dataset, with the best prediction accuracy of 97% being attained by SVM. All classifiers had an accuracy of greater than 73%. The research for the suggested work was done using a small dataset. The data that was gathered was limited to Cycles 2 and 3 of one academic year from the NAAC. Here, the QnM and QlM total scores are the only factors employed to predict grade.

In future this research can provide a deeper understanding of HEIs grade prediction by utilizing a larger dataset spanning multiple academic years and NAAC Cycles 1 to 5. In order to improve more accuracy and robustness, artificial data generation techniques such as GANs and other neural network techniques will be used in subsequent research work.

ACKNOWLEDGMENT

We would like to express our sincere gratefulness to the NAAC, which is based in Bangalore, in support of all of their assistance and guidance in assembling the data and conducting the research for this study.

References

[1] Pinto, Agostinho Sousa, et al. "How machine learning (ml) is transforming higher education: A systematic literature review." Journal of Information Systems Engineering and Management 8.2 (2023).

[2] Alyahyan, Eyman, and Dilek Duştegor. "Predicting academic success in higher education: literature review and best practices." International Journal of Educational Technology in Higher Education 17.1 (2020): 3.

[3] Cervera, Daniel Esteban Martínez, Octavio José Salcedo Parra, and Marco Antonio Aguilera Prado. "Forecasting model with machine learning in higher education ICFES exams." International Journal of Electrical and Computer Engineering 11.6 (2021): 5402.

[4] Singh, Randhir, and Saurabh Pal. "Machine learning algorithms and ensemble technique to improve prediction of students performance." International Journal of Advanced Trends in Computer Science and Engineering 9.3 (2020).

[5] Villegas-Ch, William, Milton Román-Cañizares, and Xavier Palacios-Pacheco. "Improvement of an online education model with the integration of machine learning and data analysis in an LMS." Applied Sciences 10.15 (2020): 5371.

[6] Latif, Ghazanfar, et al. "Machine Learning in Higher Education: Students' Performance Assessment considering Online Activity Logs." IEEE Access (2023).

[7] Hussain, Mushtaq, et al. "Using machine learning to predict student difficulties from learning session data." Artificial Intelligence Review 52 (2019): 381-407.

[8] Zhai, Xuesong, et al. "A Review of Artificial Intelligence (AI) in Education from 2010 to 2020." Complexity 2021 (2021): 1-18.

[9] Musthafa, M. B., et al. "Evaluation of university accreditation prediction system." IOP Conference Series: Materials Science and Engineering. Vol. 732. No. 1. IOP Publishing, 2020.

[10] Saeed, Fatek, and Anurag Dixit. "A decision support system approach for accreditation & quality assurance council at higher education institutions in Yemen." 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE). IEEE, 2015.B

[11] Nieto, Yuri, et al. "Usage of machine learning for strategic decision making at higher educational institutions." Ieee Access 7 (2019): 75007-75017.

[12] Bharti Bhojak., et al. “Investigations on Computation of CGPA for Higher Education Institutes Assessment & Accreditation by NAAC", International Journal of Emerging Technologies and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.6, Issue 5, page no. pp207-210, May-2019.

[13] Prakash, P., et al. "The Role of the National Assessment and Accreditation Council in Ensuring Quality Education in the Indian Education System: An Analysis of Its Accreditation Standards and Grading Practices." British Journal of Multidisciplinary and Advanced Studies 4.6 (2023): 1-18.

[14] Bujang, Siti Dianah Abdul, et al. "Multiclass prediction model for student grade prediction using machine learning." Ieee Access 9 (2021): 95608-95621.

[15] Attar, A. C., S. R. Kumbhar, and S. S. Kulkarni. "NAAC–Quality Benchmarking in Technical Institutions." Journal of Engineering Education Transformations. https://journaleet.in/articles/naac---quality-benchmarking-in-technical-institutions.

[16] Phyu, Thair Nu. "Survey of classification techniques in data mining." Proceedings of the international multiconference of engineers and computer scientists. Vol. 1. No. 5. Citeseer, 2009.

[17] Zhang Y. (2012) Support Vector Machine Classification Algorithm and Its Application. In: Liu C., Wang L., Yang A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg.

[18] Aria, Massimo, Corrado Cuccurullo, and Agostino Gnasso. "A comparison among interpretative proposals for Random Forests." Machine Learning with Applications 6 (2021): 100094.

[19] National Assessment and Accreditation Council (NAAC) http://naac.gov.in/images/docs/Disclosure_of_Benchmarks/Affiliated_College_Manual_20-07-2023.pdf. Accessed 02 July 2024.

[20] National Assessment and Accreditation Council (NAAC) http://naac.gov.in/images/docs/Disclosure_of_Benchmarks/Affiliated_College_Manual_20-07-2023.pdf Accessed 02 July 2024.

[21] Understanding the National Assessment and Accreditation Council (NAAC) https://www.luneblaze.com/blogs/understanding-the-national-assessment-and-accreditation-council. Accessed 02 July 2024.

[22] Grading System of NAAC http://www.naac.gov.in/docs/Grading_System.pdf. Accessed 02 July 2024.