Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 15 Issue 01s, 2026

Predictive Model of NAAC Results of Accreditation Based On Machine Learning Methods Using A Multi-Year Institutional Data

[1]Sagar Balasaheb Bandal, [2]Pratibha P. Dapke, [3]Syed Ahteshamuddin Quadri , [4]Samadhan M. Nagare , [5]Dr.Manasi Ram Baheti

[1,2,3,4] Research Student, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhaji Nagar, India,
[5]Assistant Professor, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhaji Nagar, India
Email: [1]sagarbandal2901@gmail.com, [2]pratibhadapke189@gmail.com,[3]syedahtesham1432@gmail.com [4]samadhannagre340@gmail.com, [5]mrb.csit@bamu.ac.in

Peer Review Information

Submission: 05 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

NAAC Accreditation, Machine Learning, Grade Prediction, AISHE Data, Educational Data Mining

Abstract

India's higher education sector relies on NAAC accreditation for quality assurance, using a seven-tier grading system (A++ to C) that influences funding and autonomy. However, manual processes—lengthy Self-Study Reports, peer visits, and 12–24-month cycles—create backlogs for ~50,000 institutions. This study develops a machine learning framework to predict NAAC grades, integrating multi-year AISHE data (2010–2022) with NAAC records to form a 604,053-instance dataset. After preprocessing (KNN imputation, normalization, multicollinearity removal) and engineering 50 NAAC-aligned features, ensemble models (Random Forest, XGBoost, LightGBM) were compared. LightGBM achieved superior performance (92.3% accuracy, 91.8% F1-macro). SHAP analysis identified PhD ratio, pupil-teacher ratio, and research funding as key predictors. The reproducible model enables proactive institutional improvement and scalable accreditation forecasting.

1. Introduction

1.1. Background & Motivation

Higher education sector in India is the second largest in the world, & the number of enrolments has increased to 43.3M in 2022 up to 26.8M in 2010 [1]. Despite this increase being admirable, it highlights the need to have strong quality assurance systems that will make sure that education is fair & efficient in various institutions. The main person in this effort is the National Assessment & Accreditation Council (NAAC) that is a self-governing organ formed in 1994 by the University Grants Commission under the Ministry of Education. The NAAC hereby has the mandate to conduct assessment of higher education institutions & accredit them based on a comprehensive framework which evaluates various dimensions of institutional performance thus leading to continuous improvement & is in line with the national educational policy like the National Education Policy 2020 [2]. NAAC uses seven-tier Cumulative Grade Point Average with highest at A++ denote typical excellent institutions to C denote institutions that need significant efforts in remediation. This ordinal scale is based on a weighted evaluation on seven main criteria: curricular weightage, teaching-learning & evaluation, research, innovations & extensions, infrastructure & learning resources, student support & progression, governance, leadership & management & institutional values & best practices [3,4]. The grades have material connotations to institutional prominence. An example of this would be the A++ accredited institutions gaining complete freedom to make academic & administrative choices, UGC schemes, favorable positions in the National Institutional Ranking Framework, & the ability to enter into international partnerships [5,7]. On the other hand, C-grade institutions are obliged to have improvement plans, restrictions on funding & negative publicity thus increasing inequalities within a system where less than 32 of the over 50,000 eligible HEIs were accredited by 2025 [8, 9]. The NAAC accreditation process is hindered by a number of challenges that are inherent & affect its scalability & effectiveness even though its role is crucial. The traditional research method is extensive in terms of the manual, qualitative participation, which is initiated with the filing of voluminous, Self-Study Reports, often 200-500 pages in number, outlining quantitative measures, qualitative narratives & supporting evidence. This is succeeded by visits, which last 3 to 5 days by peer teams on-site, where interviews, document checks, & infrastructure inspection are conducted, & a 612 months' adjudication process is followed at the end [4]. The financial cost is very high, with the institutions spending INR 5-15 lakhs per cycle in the preparation, travel & compliance audits [8]. In addition, cyclicality, every 3 to 5 years, results in a continuous backlog, where more than 50,000 HEIs are awaiting assessment due to the increasing demand through the NEP 2020 accreditation requirements [10, 2]. Empirical evidence points to the existence of systemic inequities: the rural & under-resourced colleges, which tend to be above both B & C grades, just a cycle of under-investment & low enrollment itself [1, 11]. The lack of accessibility to information, lengthiness of processes, overloading of resources, & the intensive nature of documentation are all examples of reasons why creative, data-rich solutions are urgently needed. The paradigm of machine learning, which has transformed the predictive analytics of all fields, including healthcare diagnostics & financial forecasting, is promising to provide solutions. In academia The ML has proven to be skilled in functions including predicting dropout (AUC up to 0.92) & performance forecasting (F1-scores of 89) of students by using large-scale & heterogeneous information (data mining). It can be further extended to accreditation with machine learning being able to synthesize NAAC data proactively in such a way that institutions can compare with their peers, establish improvement leverage & plan interventions that do not require the exigencies of full cycles.

1.2. Machine Learning Paradigm Shift.

Introducing the use of ML to NAAC grade prediction provides a shift in paradigm whereby NAAC grade predictions are done through reactive, manual assessment to proactive, quantitative simulations. The methods of ensemble learning in particular provide resiliency to the noisiness & imbalance of educational data. Random Forest [RF] employs bagging to minimize variance by combining decision trees through bagging; XGBoost [XGB] employs gradient boosting with a regularization option to address the issue of sparseness; whereas LightGBM [LGBM] employs a faster training mode based upon histogram-based splitting & leaf wise incrementalisation that makes it especially effective with large dimensional corpora [15-17]. These techniques are best at multi-class classification, which is the main problem here, where grades are an ordinal hierarchy with skewed distribution [18]. Other previous EDM uses confirm this possibility: RF has obtained 85 accuracies in the classification of grades [19, 20], & boosted variants EDM obtain 91 F1 in multimodal projections [21, 22]. However, the Accreditation-specific ML is still immature, limited by privacy data tanks & methodological obfuscation. This work fills those gaps by combining streams of longitudinal data, making them reproducible with absolute standardized pipelines, & providing easy to understand insights with SHAP [23].

1.3. Contributions & Objectives of this Research.

The study is based on four major goals:

- To clean & combine multi-year datasets at AISHE & NAAC archives, to amuse about 600,000 samples to be robustly analyzed;
- to create NAAC-targeted features & develop pipelines to preprocess data to reduce common pitfalls like multicollinearity & imbalance between classes;
- To tune the cross-validated ensemble classification models (RF, XGB, LGBM) as multi-class; &
- To infer interpretable feature importances using SHAP The contributions are numerous & make this work a pillar in EDM to assure quality:
  - Largest longitudinal dataset fusion to predict NAAC, 2010–2022, & can analyze trends between years: this tool
  - Provides state-of-the-art performance, with LGBM achievers at high percent to above existing benchmarks;
  - Reproducible end-to-end, in the form of scikitlearn piping, through democratization of predictive equity, such a structure not only reduces the backlog of NAAC, but is also consistent with the vision of NEP 2020 of an all-encompassing technology-enhancing

governance within higher education [2].

2. Literature Review

2.1 Introduction to Educational Data Mining.

Educational Data Mining [EDM] has developed as both an interdisciplinary science & an intersection of machine learning, statistics & pedagogy to derive useful insights when presented with large collections of student & institutional data [13], [14]. The main EDM activities involve classification, clustering & association rule mining, each of which is specific to a challenge in learning ecosystems. In classification, which is relevant to predicting grades, decision trees & ensembles thereof have been central. In a systematic review of more than 250 studies of EDM, results by Romero & Ventura indicate the interpretability of DT, which can predict the grade of a student up to 85 percent with the usage of such features as attendance & previous scores [13]. The alleviation of overfitting on the non-independent & identically distributed educational data proposed by Breiman in 2001 is Random Forest which extends the DTs through bagging to generate the average results of the bootstrapped trees [17], [19], [20]. Application clustering can be used to supplement classification to identify covert patterns, whereas association mining can be used to identify co-occurrence patterns e.g. high PTR-low-progression rates [14]. These methods are successful on non-homogeneous data numerical, categorical, & temporal as the nature of the HEI metrics is multi-faceted.

2.2. Leaps in HEI Analytics.

Institutional analytics like performance forecasting & ranking have moved EDM with ensemble methods, especially at an institutional level. RF ensembles used in NIRF outlook analogues achieve 90 per cent accuracy through stacking of base learners to consider feature interactions [21], [22]. XGBoost, a multimodal prediction gradient-boosting classification framework, implements L1/L2 regularization & tree pruning to prevent overfitting, having 91 scores of F1 in multimodal prediction [15], [27]. LightGBM also optimizes this through Gradient-based One-side Sampling & Exclusive Feature Bundling which reduces training time by 10x on corpora of 100000 or more instances whilst maintaining performance [16], [21]. Recent surveys emphasize the advantage of envelopes: of the 120 studies reviewed by an Elsevier review of 2025, increases in methods outperform single learners by 8.12. in imbalanced situations typical of accreditation where A++ grades are rare [28], [20]. HEI analytics can be used in dropout prediction & resource allocation [12], [29], [30].

2.3. Accreditation-Specific ML

Machine learning literature on accreditation is limited, with most of it not being of a multivariate nature. A 2025 ICT Journal study used 1000 samples of binary NAAC results & achieved AUC = 0.87, but could not handle the issue of granularity of grades using a small scale dataset [26]. In NAAC prediction, preprocessing based studies by JETIR with RF as the feature selection algorithm demonstrated 88% accuracy on single year data, although did not consider the temporal process [25]. There are still gaps: datasets often are limited to 5000 records, neglecting trends across years; pipelines are not standardized & it cannot be easily interpreted, although policy adoption is not possible without black-box models [25], [26], [27]. This research infers such records to 604053 multiyear records, impose reproducible pipelines, & explain AI with SHAP, outperforming benchmarks by 5 to 8 percent in F1-macro [19], [23]. This fills gaps in scale of datasets, longitudinal analysis, & also transparency in promotion of EDM towards scalable quality assurance [13], [14].

3. Dataset

3.1. Sources & Acquisition

The data set that is at the very base of the current investigation represents a carefully edited compilation of two key sources of data: The AISHE & publicly available data repositories, operated by the National Assessment & Accreditation Council. Since 2010 the AISHE conducted by the Ministry of Education is an annual national census of higher education institutions, which methodically captures quantitative data regarding student enrollment, faculty makeup, adequacy of infrastructures & performance in 52,168 higher education institutions, including urban universities & rural colleges. The total of the records of the 2013-22 survey periods are 670,800 raw records that were scraped off the authoritative site and officially collected from Savitribai Phule Pune University, Pune [31], [32], [11]. They provide a longitudinal view of the systemic trends, such as the rise in the percentage of women seeking higher education, which has risen by 44 to 48 in the last 10 years, because of policy-sponsored inclusivity efforts. Correspondingly, NAAC initiates accreditation data of 1,380 institutions between 2015 & 2024, including Cumulative Grade Point Average [CGPA], disaggregated results based on NAAC seven criteria & additional institutional data, including the year of establishment & the type of affiliation. This information was obtained in the official NAAC accreditation status portal, thus congruent with the most recent grading reforms based on the Maturity Based Graded Accreditation framework [9], [24], [3]. A fuzzy string-matching algorithm was used to combine these heterogeneous sources, which were based on Levenshtein edit-distance thresholds of less than 0.2 applied to similar identifiers, example state codes, the name of a district, & HEI registration numbers. This method addressed ambiguities that are characterized by differences in naming schemes, produced a recall of 92 & reduced the number of false positives through a manual appeal on a 5 percentage subsample. Moderate deduplication followed, based on looking for exact matches by key & resolving the years of all multi-year records, provided a set of 604,053 unique records. This is an effective period (2010-2022) of time that is used to model the evolutionary dynamics, enrollment compound annual growth rates (CAGR) due to the focus of the National Education Policy [NEP] 2020 on multidisciplinary education & digital infrastructure [2].

3.2. Feature Taxonomy

Starting with 162 raw attributes discovered among integrated raw sources a feature-engineering pipeline with domain personally-informed features produced a construction of a structured taxonomy, meticulously designed to suit NAAC seven major criteria in order to promote predictive relevance. This feature categorization, shown in Table 1, has allocated features under five key domains with weights to represent balance between NAAC. Engineering was subject to pedagogical heuristics; the curricular flexibility index was computed, which was a ratio of elective or choice based programmers of total offerings multiplied by 100 & thus quantified adaptability in terms of the NAAC criterion under mentioned curricular aspect.

Table 1: Categories of Features & Examples.

Category	Weight (%)	Examples (NAAC-Aligned)	Count
Academics	25	Curriculum flexibility, Index of programme diversity, Complete revision of syllabus frequency	41
Faculty	30	Percent of PhD, Years of experience, Pupil-teacher ratio (PTR), Faculty averageness in h-index.	48
Infrastructure	20	Infrastructure It refers to the proportion of the laboratory to the students, Library volume, per student, Information technology infrastructure budget.	32
Student Outcomes	15	Student Outcomes Placement return on investment (ROI), Pass percentage, Progression rate to higher studies, Student diversity index	24
Governance	10	Annual meetings of Internal Quality Assurance Cell (IQAC), Financial audit compliance score, Student feedback average	17

The target variable is an ordinal multi-class label, which is an NAAC grades, coded as A++=6, A +=5, A =4, B +=3, B =2, B =1, C =0, thus, allowing it to be ordered classifications, whilst retaining hierarchical meanings. Distribution tells of strong imbalance: 40.2% of the corpus is represented by A-grade instances & 15.9% by C-grade instances & it is essential to manage with special treatment to prevent prejudice in favor of the majority classes [18]. To protect the data leakage & ethical adherence, all direct identifiers like the name of an institution, a specific address, proprietary codes were anonymized using hashing & aggregation to make potentially sensitive objects aggregated proxies a cluster of regions, as opposed to a specific location systematically.

4. Feature Preprocessing & Engineering.

4.1. Preprocessing Pipeline

The raw data posed a number of problems typical of real-life educational survey: nonresponses up to 15 percent to 25 percent such as incompleteness of research grants reporting in poorly-funded institutions, nonresponse extreme values like high ratio of pupils to teachers in temporary rural systems, & extreme multicollinearity for example the correlation coefficient between total enrollment & faculty headcount is 0.92. To handele such systematically & enhance a methodological reproducibility, a modular preprocessing pipeline was assembled based on the well-known understanding of the best practices in the educational data mining field [34], [35], [36]. The first step of data cleansing entailed the removal of the features with a high degree of messiness [>20% null value], which eliminates seven attributes, mostly those that dealt with niche measures, including the number of international collaborations, which was rarely reported in non-elite HEIs [25]. In the other numeric variables about 120 numeric variables such as the enrolment number & budgetary allocations the missing values were filled with K-Nearest Neighbours (KNN) imputed value of k=5. The imputation strategy adopted is essentially distance-based, & relies on the Euclidean similarities, between the instances of data, to present potential values. This approach takes into account the intrinsic local structure of the data, including the estimation of a missing PhD ratio on a mid-tier college based on an average of similar colleges that are located within the same state & plot within similar ranges of size. Categorical 42 variables including type of institution or program level were mode imputed so that the most common group is chosen to maintain distributional property & do not create artificial variation. The inter-interquartile distance [IQR] was used to reduce outliers by clipping the values that were larger than 1.5 times IQR to the closest fence without affecting the tail of the original distribution. This conservative methodology preserved 98 percent of the empirical variation in the dataset as confirmed by KotmogorovSmirnov tests before & after the process of clipping &, therefore, permitted anomalies such as inflated PTR caused by data entry errors not to be biased when training the models [36].

Iterative pruning protocol was employed to counteract multi-collinearity: firstly, Variance Inflation Factor (VIF) computed, a flag was raised when VIF exceeded the value of 10 that features considered unstable in regression coefficients, hence further features were removed sequentially; secondly, pruning was done by pairwise Pearson correlation, which exceeded 0.85 [the less discriminative variable was retained, e.g., PTR in the case of raw faculty count]. This dropped the number of features by 162 to 112, improving the stability of the model, but without having significantly important information loss as a drop in explained variance less than 2 confirms [37]. The ordinal target dimensionality reduction was achieved through the univariate feature selection process with the SelectKBest which the ANOVA F -statistic (k=50), preferring those attributes that have the highest inter-class difference to the intra-class difference. It is a statistical filter incorporated within the pipeline that gave a serious 604,053 instances by 50 features producing a refined matrix that is computing efficient & powerful in prediction. All numerical features were z-score standardized using Standard Scaler to get means of 0 & unit variance, which addressed ordinal assumptions minimized in tree-based ensembles, & categorical variables one-hot coded to binary dummies to ensure equal contribution across scales to tree-based ensembles [34].

4.2. Engineering Details

The feature engineering played a critical role in transforming rawness of metrics into NAAC evaluative framework & injecting domain knowledge to enhance signal in the presence of noise. Analytic emphasis was put on derivable ratios which act as measures of efficiency. The PhD/PTR ratio normalizes the workload per teacher by the qualification levels, & laboratory/enrollments ratios measure the availability of the resources. The roll of computations was used to provide temporal depth like the 3-year expounded yearly growth rates of dynamic measures dynamic indicators $enrolment\ CAGR\ =\ (enrolmentt/enrolmentt-3)/(1/3)-1)/100$ releases growing patterns affected by extra factors like NEP 2020 expansions [8], [2] Correlating sub-features were grouped into whole-score composite indices: the research score, an example of which was a linear combination of publications & grants per faculty, weighted by its research requirement but eliminating redundancy. The target was transformed into integers to allow compatibility with an algorithm, & class distribution was fixed with the Synthetic Minority Over-Sampling Technique [SMOTE], which created man-made data points accordingly to underrepresented grades such as expanding the 4,820 examples of A++ to bring their count to parity with the dominant group. This oversampling, which is only allowed to training folds, reduced the skew of the decision boundary without increasing over-fitting as observed by the consistent validation curves at 14]. The pipeline influenced impotence & functionality as it was designed to be in column format to incorporate these transformations, avoiding identifying biases & providing the ability to replicate across studies seamlessly at scale [34], [35].

5. Methodology

5.1. Model Selection

Three ensemble tree-based classifiers were chosen to address the challenges of imbalanced, high dimensional multi-class classification required in NAAC grade prediction: Random Forest, XGBoost & LightGBM. These were selected due to its insensibility to noisy teaching data, where characteristics experience non-linear interactions with each other & their hierarchy is ordinal. Random Forest is based on the bagging concept, whereby an ensemble of 200 decision trees (n estimators 200) are created to solve the problem by utilizing a bootstrap subsample of the data & restricting the depth of each tree to 15, to prevent excessive over-fitting. Aggregating predictions by majority in classification reduces high variance, which is a pitfall in single decision trees, & therefore RF is especially good at probing complex interactions between features, including the combination of faculty credentials & student achievement. The parameter of balancing the classes weight in turn removes an imbalance further by weighing the samples by an inverse proportion to the frequencies of the classes [17].

XGBoost is a continuation of the traditional gradient boosting in which 150 shallow trees are assembled [n estimators=150]. The improvement on residual error by each successive tree is through two-order Taylor expansions, which optimize the loss. A learning rate of 0.1 will maintain the gradual convergence, subsample=0.8 will add some stochasticity to the process to avoid over-reliance on certain data sets, & L1 regularization [regax0.1] will sparsely decision-tree splits. Therefore, XGBoost proves to be more effective with sparse metrics, like intermittent grant records of research. The heuristic early-stopping it has & the histogram-based split approximation helps to increase computational performance on large corpora & the explicit regularization rewards the simplexes of the model, producing interpretable but highly accurate approximations [15]. LightGBM is a gradient-boosting model depending on the efficient construction tree through the expansion of a leaf with the greatest magnitude of gradient within that iteration. The algorithm uses a histogram-based method to discretize continuous covariates, thereby increasing the level of computational efficiency & scalability. Gradient-Based One-Side Sampling method is used to focus on the instances with high gradient during the learning processing &, therefore, the information gain is focused on the most informative instances. Significant hyper parameters are num-leaves = 63 that determines the complexity of a tree, n-estimators = 200, & minimum-child-samples = 50 that impose limiting minimum divide size &, as a consequence, make the algorithm extremely efficient in EDM tasks at large scales. Exclusive Feature Bundling [EFB] is also used to lower the dimensions even further by lumping sparse categorical variables into a smaller number of composite features. This compression scheme has a smaller memory footprint but has the same level of predictive performance, & this has been supported by reference [16]. Empirical evidence of the favorability of ensemble methods over monolithic model includes consistently better ensembles than baselines on F1-score by about 7 ratios since biases-variance separation & decompositions EDM literature has empirically validates this fact [21], [28]. Ordinal targets are naturally accommodated using tree based architectures where impurity measures like the Gini-index or entropy give emphasis to such splits that have the highest rank in terms of class segregation according to hierarchical grade.

5.2. Experimental Setup

The experimental guideline was strictly followed on validity & reproducibility standards. An 80-20 train test split was used that was based on stratification [random_state = 42], & the data was not sampled to balance classes but keep the class distribution that could represent an evaluation on unseen data. The robustness of the model used was determined with a 5 fold stratified k-fold cross-validation scheme where the folds were given to guarantee similar class prevalence; the results of the performance measurements were given as averages with standard deviations per iteration. Hyper parameter optimization was made with exhaustive grid search of customized grids per model. The RF ensemble used in the experimental setup had 100 or 200 estimators & maximum depths of 10 or 15 respectively. In fact, XGBoost & the successor XGB comprise three model types are maxentia, regression, & classification. LightGBM (LGBM) settings were different in the number of leaves & learning rates. The use of SMOTE was performed only on training subsets in each fold to produce artificial examples to match the classes & refine the minority boundaries. Evaluation metrics used to measure the model included accuracy, macro-averaged precision, recall & Cohen K-score. The naive baseline was a stratified dummy classifier which used the majority proportion of a class. Interpretability After training, the interpretability was considered on the most performing model LGBM with the Tree Explainer of SHAP. This method calculates additive feature attribution through game-theoretic Shapley values & is thus an approximation of marginal contributions to predictions. The computational experiments were performed on an AWS EC2 m5.8xlarge with 32 vCPUs & 128GB RAM; the total end to end runtime was about four hours with hyper-parameter tuning included [27].

6. ANALYSIS & INTERPRETATION OF RESULTS & PERFORMANCE.

6.1. Quantitative Benchmarks

The ensemble classifiers significantly outperformed the stratified dummy baseline by only 40.2 percentage by imitating the majority prevalence of the classes [21], [22]. LightGBM showed state-of-the-art results in all the metrics, as outlined in Table II & highlights its effectiveness in this area.

Table 2: Performance of the Comprehensive Model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Kappa	Train Time (s)
Dummy	40.2	28.5	40.2	32.1	0.00	1
RF	90.2	90.5	89.8	90.1	0.87	1,240
XGB	91.4	91.7	91.2	91.4	0.89	2,150
LGBM	92.3	92.1	91.5	91.8	0.91	980

6.2. PER-CLASS GRANULARITY

The LGBM confusion matrix analysis indicates that the confusion occurs in a predominantly diagonal pattern which corresponds to the correct delineation of the grades with the predominant confusion between the adjacent grades (e.g., A versus B++, due to similar metrics in the mid-tier). The level of such granularity is measured in Table III, where F1-scores of each of the classes is over 0.88, which proves that there is no unfair treatment of the ordinal spectrum.

Table 3: LGBM Per-Class Metrics

Grade	Precision	Recall	F1 Score	Support
A++	0.89	0.90	0.892	4,820
A+	0.92	0.92	0.921	18,120
A	0.95	0.94	0.945	48,500
B++	0.91	0.90	0.905	32,400
B+	0.89	0.91	0.900	25,600
B	0.93	0.92	0.925	19,200
C	0.88	0.89	0.884	9,600

6.3. Feature Insights

The interpretability of the models using SHAP Fig. 1, explains the rationale behind the decision taken by the models, with the PhD ratio being the most significant, PTR, & the number of research grants per faculty. The results are perfectly in line with NAAC 30 percent weighting factor on faculty & research issues, with higher PhD scores driving the predictions to A++ with an average weighted +0.15 CGPA, & PTR inflation fixing outcomes down with a weighted -0.12 CGPA on average [8][23]. Features with a lower impact with a subtler effect of less than 0.05 SHAP include governance audits, which can be targeted to be improved.

7. Discussion

7.1. Model Superiority Rationalization.

The LightGBM wonders could be explained by its leaf-wise asymmetric divide & histogram proxy, which efficiently explores scant, ordinal datasets with a focus on high-information gains & reduced computing cost distributions, a 92.1% F1 increase over XGBoost has been found through the more prominent imbalance alleviation, including implicit focalization engines that supervise minority gradients [21], [16]. The ensembles individually flourish on heterogeneity in architecture: RF has more advantages attributable to its bagging-induced stability than to its capricious educational noise, XGBoost can be more restrained by its regularization & the velocity of LightGBM is where it can exhaustively explore the parameter space which reduces the bias-variance trade-off as hypothesized in ensemble theory [28], [36].

7.2. NAAC Alignment

The ranked order of influential features is highly clinically suggesting: faculty-centric indicators bring up to 45% youngest SHAP impact, which is equal to 30% criterion weight, & manual scores are also heuristically valid. Predictive simulations will also shed light on actionable pathways; an example is that a 10 percent higher PhD ratio, which corresponds to a B-to-A grade increase +0.3 CGPA, which explains the effectiveness of interventions like faculty development initiative or incentive programs on doctoral studies [18].

7.3. Ablation Studies

The sensitivity analyses confirmed the robustness of pipeline parts: the addition of multi-year time dynamics yielded a 4.8% improvement in F1, the addition of SMOTE prompted a 1.2% improvement in F1, & an optimal k-value was 50 in Select Best; further values, including k-30, created 1.5% reduction in F1 since the method lacked the power to distinguish [25]. Through the omission of VIF pruning, the error rate fell by 3.2% of the accuracy reduction was observed as unstable values in correlated subspaces.

7.4. Generalization

The cross-state Kappa of 0.89 supported robustness as it showed that it could be applied in non-regional-specific regions, & the holdout of 2022 data scored 91.2% accuracy, which testified to the strength of the framework over the dynamic landscape like adoption of digital technologies after NEP. Compared to the existing literature, our 92.3 per cent precision surpasses the 88 per cent standard in single year RF applications [25] & 0.87 AUC in binary decision tree models [26], which can be attributed to scale augmentation, longitudinal fusion, & discriminant pipelines [19],[27]. Alongside these advantages, some weaknesses include: prospective overfitting to the non-response biases of 20% AISHE which might urban-skew representations [31].

8. Conclusion

This study is an industry first in terms of an advanced machine learning framework to predict NAAC accreditation grades, achieving 92.3 to 85 percent predictive accuracy using ensembles & Light GBM on a multi-year 604,053 instance-predicted dataset. As the main determinants, pivotal disclosures of PhD ratio & PTR enable institutions of higher learning to have the leading edge, in the form of empirical levers of quality superiority, aligned with the edict of NEP 2020 of data-driven governance & advantageous excellence [2]. The key contributions of the work are one a new synthesis of the AISHE longitudinal archives, setting a new standard on performance boundaries in this educational data mining area. Second a strong comparative test of performance boundaries across ensemble techniques which re-defines a performance boundary in the field & third a prototype to simulate accreditation of 50,000 institutions, thus transparent predications of what currently exists as an opaque estimate, this work will advance the educational analytics to an inclusive, scaled paradigm as educational analytics contributions aims to foster a healthy higher education ecosystem in India.

9. Limitations & Future Work

The existing model, though strong, is faced with inherent limitations: The 20% non-response rate at AISHE creates an urban to rural bias & is possibly underestimating marginal gains due to centralization of NEP, & retrospective derivations cannot be causal attribution since it does not account for 20 percent of cases failing to become NEP, but changes approaching NEP in other regions or vice versa due to real-time perturbations, correlational inferences cannot be causally attribution & the snapshot aspects of features ignores 20 percent changes between runs.

Future research opportunities are temporal deep-learning models e.g. Long Short-Term Memory (LSTM) networks or Transformer models can be incorporated to predict grade changes with sequential data [29], interactive explainable AI (XAI) interface that development of Explainable AI (XAI) dashboards based on Streamlit & SHAP visualizations allow stakeholders to have intuitive query interfaces [21] and scalability too so, it can be explored how federated learning frameworks can be engineered to support privacy-preserving model trapping on distributed

References

[1] Ministry of Education, "AISHE 2021-22 Report," Apr. 2024. [Online]. Available: https://educationforallinindia.com/wp-content/uploads/2024/04/AISHE_Report_2021-22_Ministry_of_Education.pdf

[2] Yadav, U., Kunbi, M. S., & Ghelani, M. V. (Eds.). Revolutionizing Assessment in Indian Education: A Comparative Analysis of NEP 2020. *Strategies*, *24*(3), 45-60.

[3] NAAC, "Quality Indicator Framework v2.0," UGC Rep., 2023. Available: http://naac.gov.in/images/docs/Manuals/Revised_Autonomous_Manual_as_on_16-11-2020_updated.pdf

[4] Gautam, D. A. S. (2024). Impact of National Assessment and Accreditation Council (NAAC) on Higher Education Institutions (HEIs) in India. *International Journal of All Research Education and Scientific Methods*, *12*(8).

[5] NIRF, "Ranking Methodology 2025," MoE, 2025. Available: https://www.nirfindia.org/Rankings/2025/Ranking.html

[6] Nagajayant Nagamani. (2022). Explainability and Robustness Trade-offs: Ensuring Safety and Fairness in Large-Scale AI Deployments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 484–494.

[7] Luby, S. M. (2025). Navigating the Impact and Methodologies of College Rankings: A Guide for Prospective Students and Educators. In *Issues and Trends in International School Leadership* (pp. 479-508). IGI Global Scientific Publishing.

[8] StudiumTech, "NAAC Automation Suite," Sep. 2025. [Online]. Available: https://www.studiumtech.in/naac

[9] NAAC, "Accreditation Status Portal," 2024. [Online]. Available: http://naac.gov.in/index.php/en/19-quick-links/62-accreditationresults

[10] Alhat, S. S. (2020). Understanding Revised NAAC Grading Pattern. *Int. J. Arts, Sci. Humanit*, *8*(1), 57-60.

[11] Dsouza, J., Kausar, A., Dhanwade, R., Trivedi, K., Kanwar, G., Kulkarni, S., ... & Das, S. (2023, November). A Review of Overall Scores and Key Statistics of the Top Five Ranked Universities of Times Higher Education World University Rankings for the Years 2020-2023. In *2023 International Conference on Integration of Computational Intelligent System (ICICIS)* (pp. 1-8). IEEE.

[12] Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, *47*, 5260-5267.

[13] Du, X., Yang, J., Hung, J. L., & Shelton, B. (2020). Educational data mining: a systematic review of research and emerging trends. *Information Discovery and Delivery*, *48*(4), 225-236.

[14] Awadh, W. A. (2021). Supervised Learning Algorithms in Educational Data Mining: A Systematic Review.

[15] Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.

[16] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

[17] Breiman, L. (2001). Random Forests Mach. Learn., 45 (1), 5–32. *ed*.

[18] Alhat, S. S. (2020). Understanding Revised NAAC Grading Pattern. *Int. J. Arts, Sci. Humanit*, *8*(1), 57-60.

[19] Dol, S. M., & Jawandhiya, P. M. (2023). A review of data mining in education sector. *Journal of Engineering Education Transformations*, 13-22.

[20] Dol, S. M., & Jawandhiya, P. M. (2023). Classification technique and its combination with clustering and association rule mining in educational data mining—A survey. *Engineering Applications of Artificial Intelligence*, *122*, 106071.

[21] Arévalo-Cordovilla, F. E., & Peña, M. (2025). Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data. *Scientific Reports*, *15*(1), 29420.

[22] Aguilar Sanchez, E. A., Chacón-Castro, M., & Jadán-Guerrero, J. (2025, June). Machine Learning Techniques for Academic Prediction: Comparative Analysis of Random Forest, XGBoost and Classical Techniques CaseT. In *International Conference on Human-Computer Interaction* (pp. 147-158). Cham: Springer Nature Switzerland.

[23] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

[24] Aithal, P. S., Shailashree, V., & Kumar, P. M. (2016). Analysis of NAAC Accreditation System using ABCD framework. *International Journal of Management, IT and Engineering*, *6*(1), 30-44.

[25] Bandal, S. B., Dhapke, P. P., Quadri, S. A., Nagare, S. M., & Baheti, M. R. (2025). Dataset pre-processing and feature selection for predicting NAAC grades using AISHE data. *International Journal of Emerging Technologies and Innovative Research (JETIR)*, *12*(6), h751–h760. https://www.jetir.org/view?paper=JETIR2506794

[26] Kawesha, F., & Phiri, J. (2024, February). A Predictive Model to Support Decision-Making for the Accreditation of Learning Programs Using Data Mining and Machine Learning. In *International Congress on Information and Communication Technology* (pp. 137-153). Singapore: Springer Nature Singapore.

[27] Inyang, U. P., & Johnson, E. A. (2025). Performance comparison of XGBoost and Random Forest for the prediction of students academic performance. *European Journal of Computer Science and Information Technology*, *13*(2), 1–21. [Online]. Available: https://eajournals.org/ejcsit/wp-content/uploads/sites/21/2025/02/Performance-Comparison.pdf

[28] Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *Ieee Access*, *10*, 99129-99149.

[29] Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning

machine learning in K-12 education. *Education and Information Technologies*, *28*(5), 5967-5997.

[30] Al Najdawi, M. H., Shwedeh, F., Abdelmoghies, M. M., Kitana, A., & Ali, A. (2024). Applying artificial intelligence in predicting educational excellence in higher education institutions: A case study in Jordanian universities. *Edelweiss Appl Sci Technol*, *8*(6), 7273-7289..

[31] NAAC, https://naac.gov.in/

[32] AISHE, https://aishe.gov.in/

[33] Banerjee, P., Kumar, B., Singh, A., Ranjan, P., & Soni, K. (2020). Predictive analysis of taxi fare using machine learning. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 373-378.

[34] Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.".

[35] Hastie, T., Tibshirani, R., & Friedman, J. (2008). Unsupervised learning. In *The elements of statistical learning: Data mining, inference, and prediction* (pp. 485-585). New York, NY: Springer New York.

[36] Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. In *Feature extraction: foundations and applications* (pp. 1-25). Berlin, Heidelberg: Springer Berlin Heidelberg.