International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

EduChoice-An Ensemble-Based Approach to College Admission Prediction

Pradnya Bachhav[1], Srushti Dhage[2], Sarthak Palde[3], Ninad Kawade[4]
[1]Assistant Professor, Computer Engineering, Pradnya Bachhav Guru Gobind Singh College of Engineering and Research Center, Nashik, Maharashtra.
[2,3,4]UG - Computer Engineering, Srushti Dhage Guru Gobind Singh College of Engineering and Research Center, Nashik, Maharashtra.
***Email ID:** pradnya.bachhav@ggsf.edu.in[1], dhagesrushti@gmail.com[2], sarthakpalde5@gmail.com[3], ninadkawade25@gmail.com[4]*

Abstract
EduChoice is a web-based application designed to assist Maharashtra CET aspirants in selecting suitable engineering colleges through machine learning–based predictions. By leveraging a hybrid approach that combines ensemble models such as AdaBoost and XGBoost, the system predicts admission chances as High, Medium, or Low based on PCM percentile, category, course preferences, and location. It features round-wise cutoff displays and advanced filtering options to enhance decision- making accuracy, transparency, and efficiency, providing a smart and user-friendly solution to simplify the college selection process.

***Keywords:** Machine Learning, AdaBoost, XGBoost, College Prediction, Ensemble Learning, Data-Driven Decision Making.*

1. Introduction
The field of engineering education in India is one of the largest in the world, with thousands of institutions offering undergraduate courses across various disciplines. Amid this extensive educational landscape, the state of Maharashtra stands out as one of the most organized, despite the complexity of its admission process. The Centralized Admission Process (CAP), managed by the Directorate of Technical Education (DTE), is a key system used to allocate admissions. Each year, hundreds of thousands of students who pass the Maharashtra Common Entrance Test (MHT-CET) must navigate a multi-layered, category-based, and round-based admission system to secure colleges that align with their academic achievements and personal interests. Although the system is well-structured, the amount of institutional data and annual changes in cutoff percentiles make the admission process a mentally demanding task for students and their families. Traditional methods for selecting colleges often rely on informal advice, paid counseling services, or manually browsing DTE portals, all of which can be time-consuming, expensive, and lack accuracy, access, and objectivity. The absence of a centralized platform that combines historical admission trends with real-time student profiles is a major gap in educational technology infrastructure. Machine learning (ML) has shown significant potential in this area over the past decade. It has demonstrated the ability to predict student performance, identify potential dropouts, and recommend personalized learning paths. Ensemble learning methods, which combine the outputs of multiple base models, are more robust and can be generalized to various learning analytics scenarios. Studies have shown that hybrid ensemble strategies, especially those using boosting algorithms like AdaBoost and XGBoost, outperform single-model classifiers when dealing with heterogeneous and multi-categorical data. In the area of admission forecasting, previous research has explored various algorithms, including logistic regression, decision trees, deep neural networks, and support vector machines. While these studies have confirmed the usefulness of ML in admission-related tasks, most have focused on general classification problems, such as graduate program admissions or international university applications. They have not adequately addressed the unique structural challenges of the Indian undergraduate engineering admission system, particularly at the state level. This gap in research inspired the development of EduChoice, a specialized, data-driven forecasting tool designed

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

specifically for Maharashtra CET aspirants. EduChoice collects student-level data, including academic performance (PCM percentile), reservation category, branch preferences, and geographical location, and compares this with a maintained database of historical DTE cutoff data. The system uses a hybrid ensemble model that combines XGBoost, AdaBoost, to generate admission probability scores—High, Medium, or Low—for each combination of college and branch. The main research goal of this work is to evaluate whether an ensemble learning model, trained in a hybrid manner, can reliably predict historical Maharashtra CET admission data and provide accurate, student-specific admission outcomes. The central hypothesis is that combining boosting-based ensemble learning with a meta-level logistic classifier will result in better predictive performance compared to any individual constituent model. The rest of this paper is organized as follows: Section II summarizes the existing literature on ML, ensemble learning, and admission prediction; Section III outlines the proposed system's approach, dataset features, preprocessing steps, and architecture; Section IV presents the experimental results and comparative analysis; and Section V concludes with a discussion of the findings, limitations, and future research directions.

2. Literature Survey

Machine learning and educational analytics have produced a rich body of research over the last two decades, with a special focus on predictive modeling of student performance. This section synthesizes relevant prior work across three thematic strands: (i) admission prediction systems based on classical and ensemble machine learning, (ii) the role of interpretability and fairness in learning algorithms, and (iii) the gap in region-specific, context-sensitive admission prediction tools.

2.1.Classical Machine Learning for Admission Prediction

Early attempts at using machine learning for admission prediction were predominantly based on single classifiers. Prince Golden et al. [9] conducted a comparative study across multiple universities, evaluating Logistic Regression, Naive Bayes, K-Nearest Neighbor, and Support Vector Machine models for predicting admission outcomes. Their results showed that while individual models achieved moderate accuracy, performance varied across institutional contexts, highlighting the sensitivity of single-model approaches to dataset characteristics.

Similarly, Assiri et al. [4] employed K-Nearest Neighbor and Decision Tree classifiers for student admission prediction in a university setting. Their study demonstrated reasonable predictive performance but also revealed limitations in handling categorical features and class imbalance. This is particularly relevant in the Indian CET ecosystem, where reservation category distributions are inherently imbalanced. Raut and Raina [8] developed a web-based college prediction application using the AdaBoost algorithm on historical DTE Maharashtra CAP data. Their work established a precedent for applying boosting-based ensemble classifiers in the Maharashtra admission context, achieving strong accuracy and demonstrating robustness in handling multi-dimensional, category-stratified cutoff datasets. This work closely aligns with the objectives of the present study.

2.2.Ensemble Learning Strategies and Their Educational Applications

The advantages of ensemble methods over single classifiers have been widely documented, both theoretically and empirically. Mohammed and Kora [7] conducted a comprehensive review of ensemble deep learning strategies, categorizing them into homogeneous methods (Bagging and Boosting) and heterogeneous methods (Stacking). Their findings indicate that heterogeneous approaches achieve better generalization when base learners are sufficiently diverse. In the domain of student performance prediction, Kumar and Bhardwaj [3] proposed a multi-level heterogeneous stacking ensemble that combines classifiers at multiple abstraction levels. Their approach achieved near state-of-the-art accuracy and reinforced the effectiveness of stacking architectures in educational classification tasks. A closely related study by Tong and Li [1] introduced a stacked ensemble model for predicting learning outcomes in Massive Open Online Courses (MOOCs). Their model incorporated multiple base learners—including KNN, Random Forest, Gradient Boosting Decision Trees (GBDT), XGBoost, and Multi-Layer Perceptron (MLP)—with

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

Logistic Regression as a meta-learner, achieving high accuracy and AUC. They also integrated SHAP (SHapley Additive exPlanations) for interpretability, demonstrating that behavioral engagement indicators such as quiz performance and forum participation were more predictive than demographic attributes. Although their context differs from admission prediction, their methodology serves as a strong reference.

2.3.Fairness, Transparency, and Interpretability

Recent research has increasingly focused on fairness and accountability in machine learning–based educational systems. Raftopoulos et al. [5] evaluated the fairness of admission prediction models using Logistic Regression, Naive Bayes, and AdaBoost. Their findings indicated that algorithmic biases—particularly those linked to socioeconomic and demographic factors—can disadvantage certain student groups. This highlights the importance of developing transparent and auditable machine learning pipelines.

Priyadarshini et al. [6] proposed an interpretable deep learning framework for undergraduate admission prediction using LIME (Local Interpretable Model-agnostic Explanations) in conjunction with Feed-Forward Neural Networks and Input Convex Neural Networks. Their work emphasizes the importance of interpretability in educational AI systems, especially for non-technical users such as students and academic counselors.

2.4.Deep Learning Approaches and Scope Limitations

Modh et al. [2] proposed a deep neural network (DNN)-based system for Maharashtra university admissions, utilizing features such as PCM percentile, family income, and domicile status. While the model achieved high prediction accuracy, deep learning approaches often suffer from reduced interpretability and increased computational complexity. This makes simpler and more interpretable ensemble methods more suitable for deployment in resource-constrained environments, such as school-level counseling systems.

2.5.Identified Research Gap

From the reviewed literature, it is evident that while machine learning–based admission prediction has been widely explored, most systems operate at a global or generic institutional level. They fail to account for the structural complexities of India's state-level CAP framework. The Maharashtra CET admission process involves category-wise, round-wise, and branch-specific cutoffs that interact in complex, non-linear ways. This creates a prediction challenge that is not adequately addressed by existing systems. Furthermore, there is a lack of end-to-end, user-friendly platforms that integrate historical DTE Maharashtra data with real-time ensemble predictions for personalized student guidance. EduChoice addresses this gap by proposing a context-specific, computationally efficient, and practically deployable hybrid ensemble framework tailored to the unique requirements of the Maharashtra CET admission ecosystem.

3. Methodology

3.1.Research Design

This paper follows a quantitative and system design–based research approach, employing a hybrid ensemble machine learning model that is developed, trained, and evaluated for its ability to classify admission outcomes of Maharashtra CET applicants. The development lifecycle follows an Agile software framework with continuous and iterative testing, feedback integration, and modular system refinement. The proposed system, EduChoice, is structured as a three-tier web application: a presentation layer (React.js), an application processing layer (Flask), and a data persistence layer (MySQL). The machine learning prediction component is implemented as a Flask-based FAST API, enabling scalable and decoupled model deployment.

3.2.Dataset Description

The dataset used in this study consists of historical admission cutoff records obtained from the official Directorate of Technical Education (DTE) Maharashtra CAP portal. Data from three consecutive academic years was collected, including round-wise, category-specific, and branch-level cutoff percentiles across multiple engineering institutes and courses in Maharashtra. The dataset includes attributes such as academic year, college name, branch/department, cutoff percentile, student reservation type (Open, OBC, SC, ST, and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

subcategories), college location (city/district), and round number (Round 1, Round 2, Round 3). Since the dataset is a curated pilot collection, data completeness and inter-year consistency were treated as key quality criteria during collection.

3.3.Data Preprocessing

Raw data obtained from the DTE portal required a multi-stage preprocessing pipeline before model training. First, records with missing or inconsistent cutoff values were identified and removed to maintain data integrity. Second, categorical variables such as student category, branch, and city were transformed using label encoding to generate numerical representations compatible with machine learning algorithms. Third, numerical features such as PCM percentile and cutoff percentile were normalized using min-max scaling to restrict values within the range [0,1], thereby reducing the effect of feature magnitude differences on model convergence. Finally, the target variable was defined as an admission probability class (High, Medium, Low) based on the relative position of the student's percentile with respect to historical cutoff ranges.

3.4.Feature Selection

The following features were identified as key predictors based on domain knowledge and dataset availability:

- PCM Percentile (aggregate score in Physics, Chemistry, and Mathematics),
- Student Category (reservation classification),
- Branch Preference (engineering discipline),
- Location/City (preferred geographic area),
- Cutoff Percentile (category- and branch-specific),
- Round Number (admission round).

These features represent the most influential variables in the Maharashtra CAP admission process, capturing both student-level academic performance and institutional admission thresholds.

3.5.Machine Learning Models

The predictive framework utilizes two models in a stacked ensemble architecture:

1. **AdaBoost (Adaptive Boosting):** AdaBoost operates by iteratively training a sequence of weak classifiers, typically shallow decision trees, where each subsequent learner focuses more on previously misclassified observations. The final prediction is obtained through a weighted aggregation of all classifiers. In EduChoice, AdaBoost is particularly effective in handling borderline admission cases where student percentiles are close to cutoff thresholds.
2. **XGBoost (Extreme Gradient Boosting):** XGBoost implements gradient boosting with a regularized objective function that reduces model complexity and prevents overfitting. It efficiently handles sparse data and missing values, making it well-suited for DTE datasets where certain category–branch combinations have limited records.

3.6.Model Training and Evaluation

The dataset was divided into training (80%) and testing (20%) sets using stratified sampling to preserve class distribution. Due to class imbalance (fewer high-probability cases), five-fold cross-validation was employed to ensure more reliable performance estimation. Evaluation metrics included Accuracy, Precision (macro-average), Recall (macro-average), and F1-Score (macro-average), ensuring a balanced assessment across all classes.

4. Results and Discussion

4.1.Results

1. Descriptive Overview

The EduChoice prediction framework was evaluated on the preprocessed DTE Maharashtra CAP dataset. Baseline performance was first measured individually for AdaBoost, XGBoost, followed by evaluation of the hybrid stacked ensemble. The results are shown in Table I.

Table 1 Comparative Model Performance

Model	Accuracy	Precision	Recall	F1-Score
AdaBoost	87.3	86.1	85.4	85.7
XGBoost	89.6	88.5	87.2	87.2
Hybrid Ensemble (Stacked)	92.4	91.8	90.7	91.2

The hybrid ensemble outperformed all individual models across all evaluation metrics, supporting the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

hypothesis that ensemble methods improve predictive performance. XGBoost emerged as the strongest individual model.

4.2.Class-Level Analysis

Class-level analysis shows that High-probability predictions achieved the highest accuracy (94.2%), indicating strong alignment between predicted and actual high-cutoff admissions. Medium-probability predictions exhibited a slight recall deficit (88.9%), possibly due to overlapping cutoff ranges. Low-probability predictions were classified with acceptable accuracy, indicating stable performance across classes.

4.3.Comparative Baseline Analysis

The proposed system was compared with prior approaches, including a standalone AdaBoost model [8], a DNN-based model [2], and homogeneous stacking ensembles. The EduChoice hybrid model demonstrated improvements of 5.1%, 3.2%, and 2.8%, respectively, highlighting the advantages of heterogeneous ensemble learning and meta-level fusion.

4.4.System Usability

In addition to predictive performance, the system was informally evaluated for usability and transparency. It generates college-wise predictions within acceptable response times and provides historical cutoff insights for better understanding. The filtering features based on fee range, location, and branch were found to be useful by users.

4.5.Limitations

This study has several limitations. First, the dataset is limited in size despite being sourced from an official government portal, which may affect generalization. Second, predictions are based on historical data and cannot account for sudden policy or seat allocation changes without retraining. Third, the system currently focuses only on Maharashtra CET admissions and has not been validated for other examinations such as JEE or NEET. Finally, advanced explainability features such as SHAP-based visualizations have not yet been implemented, although they are planned for future work.

Conclusion

This paper gave an example of EduChoice as a hybrid ensemble machine. learning system that was created to aid data-driven college selection of Maharashtra CET engineering applicants. By combining AdaBoost and XGBoost in an architecture consisting of stacked models with use of. The system uses a Hybrid Ensemble in order to achieve a precision of 92.4 in predicting the admission rates between categories. It has a three-tier web architecture which ensures scalability and accessibility, whereas such functionalities as historical cutoff visualization increase transparency and decision making. The study emphasizes the excellence of group learning and domain- accurate and efficient as well as efficient building of the features in the engineering have useful educational prediction systems without the use of deep learning models are computationally intensive.

Future Scope

The EduChoice system will be improved by emphasizing further developments in the future on increasing the datasets to more years and broader. The memes are geographically generalized. Integration real-time DTE portal data using APIs will be possible. Updates in real time and retraining of the model. Additionally, it can be used to support other competitive exams with the system including JEE and NEET, extending its way of applicability and influence to a greater number of students in India

References

[1]. T. Tong and Z. Li, “Predicting learning achievement using ensemble learning with result explanation,” PLoS ONE, vol. 20, no. 1, e0312124, Jan. 2025, doi: 10.1371/journal.pone.0312124.

[2]. H. Modh, M. Bharati, A. Mamgai, and H. Tiwari, “A University Admission Prediction System Using Machine Learning and Deep Neural Networks,” IJCRT, vol. 12, no. 4, Apr. 2024.

[3]. M. Kumar and V. Bhardwaj, “Ensemble Learning Based Model for Student’s Academic Performance Prediction,” Ing ´enierie des Syst `emes

[4]. d’Information, Oct. 2024. B. Assiri, M. Bashraheel, and A. Alsuri, “Enhanced Student Admission Procedures Using Data Mining and ML Techniques,” Applied Sciences,Jan. 2024.

[5]. G. Raftopoulos, G. Davrazos, and S. Kotsiantis, “Fair and Transparent Student

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3436-3441
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0444

Admission Prediction Using Machine Learning Models,"

[6]. Algorithms, Dec. 2024. A. Priyadarshini, B. Martinez-Neda, and S. Gago-Masague, "Admission Prediction in Undergraduate Applications: An Interpretable Deep

[7]. Learning Approach," Elsevier, Jan. 2024. A. Mohammed and R. Kora, "A Comprehensive Review on Ensemble Deep Learning," Journal of King Saud University – Computer and Information Sciences, Feb. 2023.

[8]. N. Raut and S. Raina, "College Prediction System," IJSREM, Nov. 2023.

[9]. P. Golden et al., "A Comparative Study on University Admission Predictions Using ML Techniques," IJSRCS Engineering & IT, Mar.–Apr. 2021.

[10]. K. Kumari, M. Kataria, V. Limbani, and R. Soni, "CAPSLG: College Admission Predictor and Smart List Generator," in Proc. ICAST, 2019