International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-8

Cutoff Trend Prediction for Engineering Admissions Using Machine Learning

Mule Darshan Dipak, Bhosale Sachin Bajirao, Bangar Abhishek Popat
Dept. of Computer Engineering,
Jaihind collage of Engineering, Pune,India

Abstract. Engineering admission in Maharashtra is carried out through merit-based Centralised Admission Process (CAP) rounds, in which every college, branch and reservation category closes at a different cutoff percentile each year. Students usually rely on the previous year's cutoff list to decide their preference order, but cutoffs move from year to year because of changes in seat matrices, examination difficulty, branch demand and category-wise competition. Existing college-predictor tools mostly compare a candidate's score with last year's closing val-ues and rarely model how the cutoff itself is changing. This paper reviews five selected studies on admission prediction, cutoff-based college prediction and time-series forecasting, and iden-tifies gaps in temporal modelling, validation and uncertainty reporting. Based on these findings, a Machine Learning Framework for Cutoff Trend Prediction is proposed. The framework con-verts official CAP cutoff lists into a structured dataset of about 239,840 records, derives lag-based and trend features for every college-branch-category-round combination, and trains re-gression models such as Ridge regression, Random Forest and gradient boosting, with LSTM and ARIMA considered as optional time-series baselines. The forecast cutoff for the coming year is compared with a student's percentile to group colleges as Safe, Target or Dream. A pro-totype web application with search, prediction and history modules has been devel-oped to de-liver the results. The framework aims to improve the accuracy, transparency and practical use-fulness of cutoff information for engineering aspirants. Quantitative evaluation of the proposed models is identified as future work.

Keywords: Cutoff Prediction; Engineering Admissions; Machine Learning; Time-Series Forecasting; Random Forest; Gradi-ent Boosting; LSTM; College Predictor; CAP Rounds; Web Application

I. Introduction

Admission to undergraduate engineering programmes in India is highly competitive. In Maharashtra, seats in government, aided, university and private colleges are allotted through the Centralised Admission Process (CAP), where candidates are ranked by their entrance-test percentile and receive seats round by round. Every combination of college, branch and reservation category closes at a different cutoff percentile in each round, and these values are published as long, table-heavy PDF docu-ments. A candidate must prepare a preference list before the first round, and a poorly ordered list can waste a good percentile or leave the candidate with an unwanted seat.

The common practice is to read the cutoff of the previous year and assume that the new year will behave in a similar way. Online and desktop college-predictor tools fol-low the same idea and filter colleges according to the student's score, category and re-gion [2]. However, a closing percentile is not fixed. It changes with the number of ap-plicants, the difficulty of the

examination, the seat matrix, the popularity of a branch and the competition within each category. A tool that only looks up last year's value therefore gives no information about the direction or speed of the change.

Machine learning has been applied to several admission-related problems. Most studies predict whether an individual applicant will be admitted, using academic scores and background attributes [1], [3], [4]. A different line of work forecasts time series with statistical models such as ARIMA and deep recurrent models such as LSTM [5], [8], and tree-based ensembles are widely used for tabular prediction [6], [7]. These ideas have not been combined into a framework that forecasts the cutoff trend of each college-branch-category combination from official counselling data.

This study makes four contributions. First, it reviews five selected papers and iden-tifies the gaps that matter for cutoff fore-casting. Second, it describes a structured dataset of about 239,840 cutoff records extracted from official CAP cutoff lists. Third, it pro-poses a machine learning framework that uses lag, trend and categorical features to predict the cutoff percentile of the next year and converts the forecast into Safe, Target and Dream groups. Fourth, it presents a prototype web application that delivers search, prediction and history features to students.

II. Review Methodology

A systematic review of five selected research papers was carried out to understand cur-rent work on admission prediction and forecasting. The papers were identified through keyword searches on topics such as engineering admission prediction, college predic-tor, cutoff-based college prediction and time-series forecasting. Studies were selected when they dealt with admission outcomes, cutoff-based college recommendation or forecasting methods that can be adapted to cutoff trends, and when their objective, data and findings could be verified from the published record. The selected studies were compared on their objective, data, method, major findings and limitations [1]-[5]. Four further references [6]-[9] were used only to describe the algorithms and forecasting concepts adopted in the proposed framework.

The review shows that random forest and other ensemble classifiers are popular for predicting admission outcomes [1], [2], while regression and classification comparisons are common for graduate admissions [3], [4]. Forecasting studies show that recurrent neural networks can reduce error compared with ARIMA on long economic series [5]. Only one of the five studies deals directly with cutoff data of Maharashtra engineering colleges [2], and none of them forecasts how the cutoff of a college-branch-category combination changes from year to year. This gap motivates the framework proposed in this paper.

III. Literature Review

1. Admission Outcome Prediction Using Ensemble Learning

Gupta et al. [1] studied automated undergraduate engineering admission prediction for an Indian university in the northern region. Their dataset contained more than 1,400 student records, and the attributes included secondary and senior secondary percent-ages, preparation method, family background, school board type, drop cases and finan-cial aid. Several classification methods were compared, and a random forest classifier achieved an accuracy of 87% on the chosen dataset. The authors argue that such a model can support students and parents in early college selection and can help universities plan corrective measures. The study shows that ensemble classifiers handle mixed academic and background attributes well. However, it works at the level of individual applicants in a single uni-versity. It does not use the published cutoff lists of a counselling system and therefore cannot say how closing percentiles will change in the following year.

2. Cutoff-Based College Prediction for Maharashtra

Dhavale et al. [2] presented Academic Matcher, a Windows application that lets stu-dents look up previous-year CET cutoffs of engineering colleges across Maharashtra. A student enters the CET score, caste category and region preference, and the applica-tion returns a short list of colleges where admission is likely. Support vector machine, decision tree and random forest algorithms are named as the prediction techniques, and the results section shows sample output with five suggested colleges.

This work is the closest to the present study because it targets the same counselling context. The paper reports no quantitative evaluation such as error or accuracy values, it treats the previous cutoff as the basis of the prediction, and it is delivered as a desktop application. A forecast of the new cutoff, measured against held-out years, is therefore still missing.

3. Regression and Classification for Graduate Admissions

Acharya et al. [3] compared several regression models for predicting the chance of ad-mission to graduate programmes, using applicant academic parameters collected from an Indian perspective. The comparison of linear, support vector, decision tree and ran-dom forest regression made the study a common reference for later work on admission chance estimation. Iman and Tian [4] compared Naive Bayes, logistic regression, a multilayer perceptron and decision trees on a dataset of 400 applicant records and found Naive Bayes to be the most accurate on that data. They also proposed a framework in which machine learning classification supports the admission decision process.

Both studies show that simple, well-chosen models can work on small admission datasets, and that model comparison with cross-validation is an accepted practice. Their limitations are the small number of records, the focus on postgraduate admission and the use of applicant attributes instead of institution-level cutoff history.

4. Time-Series Forecasting and Ensemble Learning

Siami-Namini and Siami Namin [5] compared ARIMA with LSTM for forecasting eco-nomic and financial time series and reported that the LSTM model reduced the average error by about 84 to 87 percent relative to ARIMA. Hochreiter and Schmid-huber [8] introduced the LSTM architecture, which can learn long-range dependencies in se-quences. Classical forecasting texts such as Hyndman and Athanasopoulos [9] recom-mend simple benchmarks, namely the naive forecast and the moving average, before more complex models are accepted.

For tabular data with engineered lag features, Random Forest [6] and gradient boost-ing as implemented in XGBoost [7] are strong and widely used choices. A cutoff series is short, with one value per admission year for each college-branch-category-round combination, whereas the studies on LSTM and ARIMA used long series. It is therefore reasonable to compare lag-feature ensembles with LSTM and ARIMA instead of as-suming that a deep model will be best.

Overall, the reviewed studies show that ensemble methods work well for admission data, that regression and classification comparisons are useful, and that forecasting models can be adapted to sequences [1]-[5]. They also show that existing work concen-trates on applicant-level outcomes or on look-up of last year's cutoff. An integrated framework that forecasts cutoff trends from large counselling datasets and reports the result with clear uncertainty bands is still needed.

IV. Comparative Analysis

The reviewed studies differ in their objective, data and method. Their comparison is presented in Table 1.

Table 1. Comparison of the reviewed studies.

Author, Year	Objective	Method / Technology	Major Reported Findings	Limitations
Gupta et al., 2025 [1]	Engineering admission pre-diction	Random forest and other classifi-ers on student attributes	Random forest reached 87% accu-racy on more than 1,400 records	Single university; applicant-level data; no cutoff trends
Dhavale et al., 2024 [2]	College pre-diction from CET cutoffs in Maharashtra	SVM, decision tree, random for-est; Windows application	Suggests five likely colleges from score, category and region	No quantitative evaluation re-ported; uses previous cutoff only
Acharya et al., 2019 [3]	Graduate admission chance estimation	Comparison of regression models	Shows regression comparison is practi-cal for admission chance	Postgraduate applicants; small dataset
Iman and Tian, 2021 [4]	Graduate admission deci-sion prediction	Naive Bayes, logistic regression, MLP, decision tree	Naive Bayes was most accurate on 400 records	Small dataset; no temporal analysis
Siami-Namini and Siami Namin, 2018 [5]	Forecasting economic and financial series	ARIMA compared with LSTM	LSTM reduced aver-age error by 84 to 87% over ARIMA	Long financial series; not tested on admission data

V. Proposed Conceptual Architecture

1. Dataset

The proposed framework is built on a structured dataset of about 239,840 cutoff records extracted from official CAP cutoff lists and stored as a CSV file. Each record describes the closing cutoff percentile of one college, branch and category in a given year, round and stage. The attributes are listed in Table 2. The cutoff percentile is the prediction target, and the last two attributes are kept only to trace every record back to its source document.

Table 2. Attributes of the cutoff dataset.

Attribute	Description	Role
year	Admission year of the cutoff list	Feature / time index
college	Name of the engineering institute	Feature
branch	Engineering programme or branch	Feature
category	Reservation or seat category	Feature
round	CAP round number	Feature
stage	Allotment stage within the round	Feature
cutoff_percentile	Closing percentile of the seat	Target variable
choice_code	Code identifying the institute and branch choice	Identifier
pdf_line, source_pdf	Line number and name of the source PDF	Traceability only

International Journal of Statistics and Mathematical Research Trends
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-8

2. Architecture Flow

The architecture is shown in Fig. 1. Layer 1 collects the round-wise cutoff lists pub-lished by the admission authority for each year. Layer 2 extracts the tables from the PDF files into CSV format, removes duplicate and malformed rows, and normalises college and branch names so that the same institute is not stored under several spellings. Layer 3 stores the cleaned records in a PostgreSQL database with indexes on college, branch and category for fast search. Layer 4 derives features for every college-branch-category-round series. Layer 5 trains and compares the forecasting models. Layer 6 validates them with a time-based split, and Layer 7 produces the cutoff forecast that is converted into Safe, Target and Dream groups and shown in the web application.

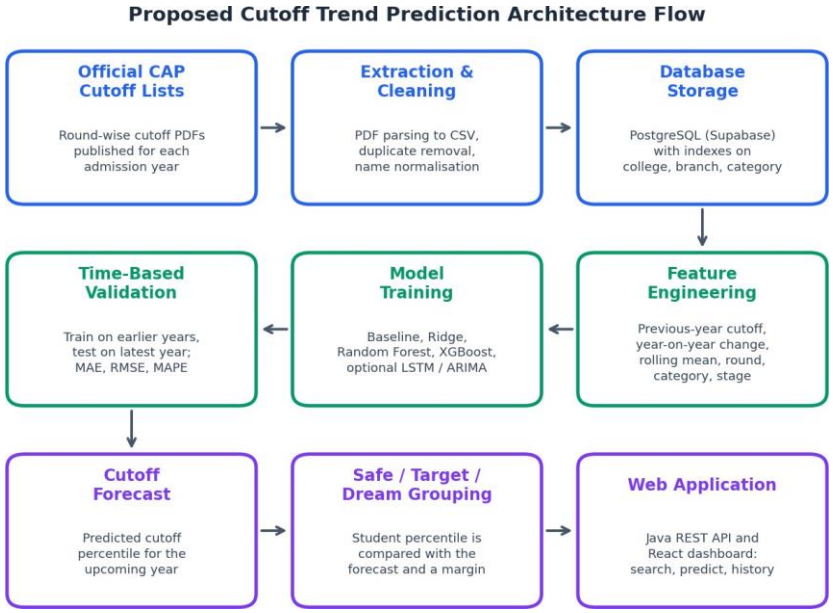


Fig. 1. Proposed cutoff trend prediction architecture flow.

3. Feature Engineering and Models

For each college-branch-category-round series, the cutoff of year t is modelled as a function of earlier years and of the series identity, as shown in (1). The year-on-year change in (2) captures the direction and speed of the trend. Rolling means over the last two or three years, the round, the stage and the encoded category and branch are added as further features.

$$c_t = f\left(c_{t-1}, c_{t-2}, \Delta c_{t-1}, x \right) \quad (1)$$

$$\Delta ct\text{-}1 = ct\text{-}1 - ct\text{-}2 \, (2)$$

Here ct is the cutoff percentile in year t and x is the vector of categorical and calendar features. The simplest benchmarks are the naive forecast, which repeats the previous cutoff, and the moving average [9]. These are followed by Ridge regression, Random Forest [6] and XGBoost gradient boosting [7]. An LSTM [8] and an ARIMA model are included only for series that are long enough, so that the effect of model complexity can be measured instead of assumed [5].

4. Evaluation Protocol

Random shuffling would leak future information into training, so a time-based split is proposed. Models are trained on all years up to year T-1 and tested on year T, and the procedure is repeated with each available year as the test year. Forecast quality is measured with the mean absolute error in (3) and the root mean squared error in (4), where yi is the actual and ŷi the predicted cutoff percentile of the i-th test record. Mean abso-lute percentage error and the R2 score are reported as supporting measures.

$$MAE = (1/n)\ \Sigma\ |\ y_i - \hat{y}_i\ |\quad (3)$$

$$RMSE = \sqrt{(\ (1/n)\ \Sigma\ (\ y_i - \hat{y}_i\ )^2\ )}\ (4)$$

A model is accepted only if it beats the naive and moving-average benchmarks on the held-out years. The forecast is then converted into a recommendation by comparing it with the student's percentile p, as in (5). With two configurable margins $\delta 1$ and $\delta 2$, a college is grouped as Safe when $d \leq -\delta 1$, as Target when $-\delta 1 < d \leq \delta 1$, and as Dream when $\delta 1 < d \leq \delta 2$. The grouping can be checked against the seats that candidates with similar percentiles actually received in earlier years.

$$d = \hat{c}_t - p(5)$$

5. Prototype Implementation

A prototype of the system, named CollegeCutoffAI, has been developed. The cutoff records are stored in a PostgreSQL data-base hosted on Supabase. A Java backend ex-poses a REST interface that answers college and branch search queries from the data-base, and a React front end provides the home page, college search, cutoff prediction and prediction history pages. Be-cause the interface is a web application, it can be used on both desktops and mobile phones, which addresses the platform limitation of earlier desktop tools [2]. Integration of the trained forecasting models with the prediction page is the next imple-mentation step.

VI. Research Gaps and Future Directions

The gaps identified from the reviewed literature and the corresponding future directions are summarised in Table 3.

Table 3. Research gaps and future directions.

Gap supported by evidence	Supporting studies	Future direction
No reviewed study forecasts how the cutoff of a college-branch-cate-gory combination changes over the years	[1]-[5]	Forecast next-year cutoff percentile using lag and trend features
Admission studies use appli-cant-level attributes and small da-tasets	[1], [3], [4]	Use large institution-level counselling data of about 239,840 records
Cutoff-based predictors rely on last year's value and report no quantitative evaluation	[2]	Validate with a time-based split using MAE, RMSE and benchmark comparison

Time-series models were tested on long financial series, while cutoff series are short	[5]	Compare lag-feature ensembles with LSTM and ARIMA on short annual series
Round, stage and category ef-fects are rarely modelled to-gether	[2]	Include round, stage, category and branch as model features
Uncertainty is not communi-cated to students	[1], [2]	Report Safe, Target and Dream groups with configurable margins
Existing predictor tool is a desktop application	[2]	Deliver a responsive web application for desktop and mobile users

Directions that still require investigation include prediction intervals around each fore-cast, handling of new colleges and branches that have no history, the effect of seat-matrix changes on cutoffs, expansion of the dataset to more admission years, and com-parison of the forecast with the actual cutoffs once the next admission year is published.

VII. Conclusion

This paper reviewed five studies on admission prediction, cutoff-based college predic-tion and time-series forecasting. The review shows that ensemble models are effective for admission data and that forecasting methods can be adapted to sequences [1]-[5]. It also shows that existing work either predicts applicant-level outcomes or looks up last year's cutoff, and that the year-to-year movement of cutoffs is not modelled.

To address this gap, the paper proposed a machine learning framework for cutoff trend prediction. The framework uses a structured dataset of about 239,840 CAP cutoff records, lag and trend features, a comparison of Ridge regression, Random Forest, XGBoost and optional LSTM and ARIMA models against naive benchmarks, and a Safe, Target and Dream grouping based on the forecast. A prototype web application has been built to deliver the results to students.

The proposed framework provides a direction toward more transparent and data-driven college selection. Training the models on the full dataset, evaluating them on held-out admission years and integrating the best model into the web application will be required to validate its effectiveness and practical usefulness.

References

1. M. Gupta, Alpana, P. Gupta, and N. Varshney, “A comparative study of automated undergraduate engineering admission prediction in an Indian university using ma-chine learning,” Journal of Computational Social Science, vol. 8, no. 3, pp. 1–22, 2025. DOI: 10.1007/s42001-025-00384-w
2. R. Dhavale, P. Ugale, A. Hanwate, and M. Umare, “Academic Matcher – The Col-lege Predictor,” International Journal for Research in Applied Science & Engi-neering Technology (IJRASET), vol. 12, no. IV, 2024. DOI: 10.22214/ijra-set.2024.60617
3. M. S. Acharya, A. Armaan, and A. S. Antony, “A Comparison of Regression Mod-els for Prediction of Graduate Admis-sions,” in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019.
4. A. Iman and X. Tian, “A Comparison of Classification Models in Predicting Grad-uate Admission Decision,” Journal of Higher Education Theory and Practice, vol. 21, no. 7, 2021. DOI: 10.33423/jhetp.v21i7.4498

5. S. Siami-Namini and A. Siami Namin, “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” arXiv preprint arXiv:1803.06386, 2018. DOI: 10.48550/arXiv.1803.06386
6. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324
7. T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Pro-ceedings of the 22nd ACM SIGKDD International Conference on Knowledge Dis-covery and Data Mining, pp. 785–794, 2016. DOI: 10.1145/2939672.2939785
8. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computa-tion, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735
9. R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021.