INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 07 | JULY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

Implementation paper on "Admission Sarthi" - College Predictor

[1]Saurav Shrikant Bansode , [2]Prathamesh S. Mahindrakar , [3]Shreyash Kadam , [4]Snehal P. Hon
[1,2,3] U.G. Student, Department of Electronics and Computer Engineering , P.E.S Modern College of Engineering,
[4]Assistant Professor, Department of Electronics and Computer Engineering,
P.E.S Modern College of Engineering,
Pune, Maharashtra, India
saurav_bansode@moderncoe.edu.in, prathamesh_mahindrakar@moderncoe.edu.in, shreyash_kadam@moderncoe.edu.in,
snehal.hon@moderncoe.edu.in

***Abstract*— Selecting the right college is crucial for shaping students' future careers. Unfortunately, many students, especially in rural areas, end up in unsuitable colleges due to limited information. "Admission Sarthi" - College Predictor addresses this issue by focusing on Maharashtra's Engineering Colleges and MHT-CET scores. It provides a user-friendly platform where students can enter their marks or rank and receive a list of potential colleges along with admission probabilities. Using real and official data, the tool ensures accurate and reliable predictions. Admission Sarthi features an AI bot to help students with questions related to colleges, branches, and cutoffs, enhancing their decision-making process. Additionally, students can export the generated college list to PDF and Excel formats. The platform includes mock tests to aid in exam preparation. The goal is to guide students towards informed decisions, helping them avoid choices that could negatively impact their academic and professional paths.**

Keywords—Machine Learning, Cloud Computing, Admission.

I. INTRODUCTION

The journey to higher education often begins with a critical decision which is selecting the right college. This decision can be a turning point, significantly impacting a student's future career trajectory. However, choosing the ideal institution can be a perplexing and stressful endeavor, further complicated by the limited access to reliable and comprehensive information about available colleges.

Across the educational landscape, it's an unfortunate reality that students, despite having promising academic records and higher ranks, may find themselves securing admission to institutions that may not fully harness their potential. This disheartening scenario often arises due to a lack of awareness about suitable colleges, coupled with the pernicious influence of misleading, paid rankings.

"Admission Sarthi" addresses this challenge. The project demystifies the complex admission process, particularly for Maharashtra Engineering Colleges, by leveraging the MHT-CET cutoffs/scores. The platform offers students a user-friendly interface where they can input their marks or rank, and in return, receive a meticulously curated list of colleges they are most likely to secure admission in. The distinguishing feature is the addition of probability percentages, which further clarify the likelihood of admission.

In addition to providing accurate and reliable predictions, Admission Sarthi enhances the student experience. Users can download the generated college list in Excel and PDF formats, allowing them to review and analyze their options offline. This feature is particularly useful for students who may not have continuous internet access or prefer to have a hard copy for reference.

Furthermore, as part of the commitment to comprehensive support, the project integrates mock tests into the platform. These mock tests simulate real exam conditions, providing students with a valuable tool to assess their readiness and identify areas for improvement. Through these mock tests, students can gain confidence and refine their preparation strategies, ultimately increasing their chances of success in entrance exams.

At the core of Admission Sarthi is the dedication to providing students with information derived from real and official data, ensuring the highest degree of accuracy in predictions and recommendations. The mission is to empower students with knowledge, enabling them to make informed decisions that align with their academic potential and personal preferences.

As part of the endeavor to assist students holistically, the project incorporates advanced technologies like Generative AI and OpenAI-powered bots. These technologies are seamlessly integrated into the platform, providing students with immediate access to reliable information and personalized assistance whenever they need it. By harnessing the power of AI, Admission Sarthi aims to revolutionize the college selection process, making it more transparent, efficient, and tailored to each student's unique needs.

Implementing this college predictor tool aims to eliminate uncertainties, guide students toward their best-fit institutions, and significantly bolster their confidence and success in the dynamic field of higher education.

II. LITERATURE SURVEY

Fathiya and Sadath (2021) [1] proposed an "University Admissions Predictor Using Logistic Regression" that leverages machine learning and data collection through web scraping to aid prospective graduate students in evaluating their chances of admission into specific universities. The admissions predictor employs logistic regression as its classification algorithm, chosen for its simplicity and efficiency, making it a practical choice for training models. Users interact with the predictor by inputting their GRE test scores and GPA, and the system generates admission probability estimates. The research underscores the advantages of web scraping for data acquisition, noting an 88% reduction in time compared to manual data collection methods. Additionally, data visualization is employed to enhance user understanding. The admissions predictor proves valuable for prospective graduate students, helping them make informed decisions about where to apply based on their likelihood of acceptance.

Sivasangari et al. (2021) [2] recognized the challenge students face when choosing the right university for higher studies. Existing advisory services and online applications are often costly and inaccurate. The study aimed to create a predictive model using linear regression, random forest, and CatBoost algorithms to accurately estimate the probability of university admission. The research stressed the importance of higher education in shaping students' futures, particularly for those aspiring to study abroad. The model considered parameters such as GRE Score, TOEFL Score, University Ranking, Proposal Statement, Recommendation Letter Power, Undergraduate GPA, and Study Experience to predict admission likelihood. The work involved dataset division, preprocessing, and flexible criteria based on university cutoffs. This study contributes valuable insights for students seeking higher education opportunities.

Goni et al. (2020) [3] addressed the challenges of graduate admission selection, where universities employ diverse criteria, including GRE scores, CGPA, and more. Existing web applications and consultancy services often lack accuracy. To address this, they introduced a Deep Neural Network (DNN) model for comprehensive admission prediction, incorporating all criteria. The DNN model outperformed existing methods with a promising R-squared score of 0.8538 and an MSE of 0.0031. Recognizing the complexity of graduate admissions, including exams and portfolio assessment, Goni et al. emphasized the need for an efficient and accurate guidance system. Their proposed DNN-based system provides a solution that surpasses existing approaches, offering students a more reliable way to select universities and enhance their chances of admission.

Sridhar, Mootha, and Kolagati (2020) [4] addressed the challenge of graduate students selecting universities for admission, where high application costs and competitive admission rates often leave students uncertain about which institutions align with their profiles. They introduced a stacked ensemble model for university admission prediction, taking into account various factors, including research and industry experience. Compared to previous prediction systems, their model demonstrated superior accuracy, outperforming other machine learning algorithms with an impressive 91% accuracy. The study emphasized the importance of guiding students in university selection and suggested future improvements, such as considering Statement of Purpose essays and Letters of Recommendation to enhance prediction accuracy and applying Natural Language Processing for evaluation.

In their "MBBS College Predictor," Tippana et al. (2023) [5] address the complex process of selecting the right MBBS college. Their system predicts the likelihood of student admission to specific colleges based on available seats, the student's area, category, and grades, simplifying college selection and automating the process. Additionally, the predictor incorporates the All India Quota, making it more comprehensive and beneficial for students from various regions. This practical solution streamlines university choices during the admissions process, providing students with valuable insights and guidance in navigating the competitive landscape of MBBS admissions.

In their study on "Engineering & Technology Admission Analysis and Prediction," Bhoite and More (2020) [6] focus on guiding students in making informed decisions about engineering college admissions after completing their Higher Secondary Schooling. They employ various machine learning algorithms, including Logistic Regression, K Nearest Neighbors, Decision Tree Classifier, Random Forest Classifier, Naive Bayes, and Support Vector Machine, to create a predictive model that helps students determine their admissibility and suggests suitable colleges. By evaluating and comparing the performance of these classifiers, the authors find that Random Forest and Decision Tree classifiers yield the best accuracy, contributing to a more informed and automated decision-making process for students.

Wang and Jia's study, titled "Research on Prediction of College Students' Performance Based on Support Vector Machine" [7], delves into leveraging college entrance examination results to forecast university course performance. This predictive model serves to empower students with informed decision-making capabilities while aiding colleges in refining their teaching standards. The research harnesses the power of a support vector machine to construct a predictive model for college course performance derived from entrance exam outcomes. Through meticulous cross-validation techniques to optimize model parameters, the study achieved an impressive prediction accuracy rate of 73.6%, showcasing the efficacy of machine learning in enhancing educational outcomes..

The study authored by D.A. Chithra [8] and colleagues focuses on developing a machine learning model called UAP (University Admission Predictor) for predicting a student's chances of gaining admission to universities in the United States. This model aims to help students make informed decisions about their higher education options and can consider multiple crucial factors affecting the admission process. This research contributes significantly to the field of educational technology by leveraging machine learning techniques to address the complexities of college admissions,

paving the way for more accurate and personalized guidance for students worldwide.

Zhenru Wang and Yijie Shi [9] employed machine learning, specifically the Adaboost algorithm, to forecast college admission thresholds using college entrance examination (CEE) outcomes. This ensemble learning technique showcased remarkable performance improvements over conventional methods. By incorporating variables like enrollment projections, applicant volumes, and test question complexities, the Adaboost algorithm delivered highly precise predictions of admission thresholds. Presented at the 2016 IEEE International Conference on Computer and Communications, this study concluded with experimental validation of the model's accuracy, providing a valuable reference point for students preparing for the college entrance examination.

Amal AlGhamdi, Amal Barsheed, Hanadi AlMshjary, and Hanan AlGhamdi [10] introduced a machine learning approach for predicting graduate admission in their paper presented at the 2020 2nd International Conference on Image, Video, and Signal Processing. The authors addressed the challenges faced by graduates in identifying suitable universities for postgraduate studies. The traditional approach of seeking advice from consultancy organizations was deemed biased and inaccurate. To mitigate this, the paper proposed a machine learning approach to predict postgraduate admission, evaluating three regression learning strategies: linear regression, decision tree, and logistic regression. Logistic regression demonstrated the highest accuracy, making it the suggested model for predicting future applicants' chances of university admission.

Thomas Lux, Randall Pittman, Maya Shende, and Anil Shende [11] delved into the application of supervised learning techniques on undergraduate admissions data in their work presented at the ACM International Conference on Computing Frontiers (CF '16). The authors addressed the challenge faced by colleges and universities in efficiently utilizing the wealth of data gathered over the years to guide admissions decisions. They proposed the use of supervised learning techniques, specifically perceptrons and support vector machines, to predict admission decisions and enrollment based on historical applicant data. Experimental results demonstrated that a classifier, trained on past data, could accurately identify likely accepted applicants and those likely to enroll. Feature selection experiments provided insights into the significance of applicant features, aiding future data collection and decision-making by admissions offices.

M. S. Acharya, A. Armaan, and A. S. Antony [12] explored the realm of graduate admissions prediction using machine learning in their paper presented at the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS). The authors addressed the common dilemma faced by prospective graduate students in choosing universities for master's programs and the unreliability of existing predictors and consultancies. They proposed a machine learning-based method employing regression algorithms, including Linear Regression, Support Vector Regression, Decision Trees, and Random Forest, to predict graduate admissions. The paper detailed the methodology, including the evaluation of the dataset using various regression methods, such as Linear Regression, Support Vector Regression, Decision Tree, and Random Forest. The authors concluded by comparing the performances of these models based on Mean Squared Error (MSE) and R2 Scores.

III. BLOCK DIAGRAM

Below is the diagram illustrating the Continuous Integration and Continuous Deployment (CI/CD) workflow for the Admission Sarthi - College Predictor system:

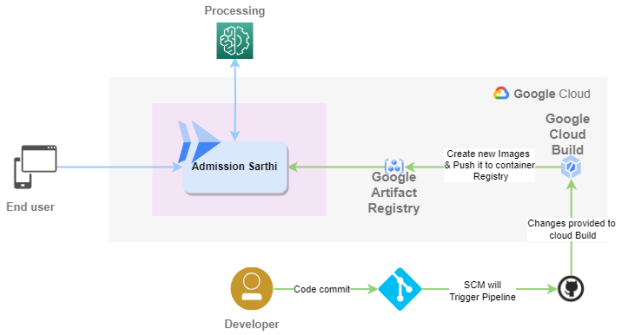

Fig 1. Google Cloud Deployment

F.1 Developer Commits Code: The developer pushes code changes to the backend repository, which is typically hosted on a version control system like Git.

F.2 Source Control Management (SCM) Trigger: The SCM (e.g., GitHub) triggers a pipeline upon detecting new code commits or changes.

F.3 Cloud Build Integration: The changes are provided to Google Cloud Build, which is configured to automatically build and package the application based on the code changes.

F.4 Container Image Creation: Google Cloud Build creates a new container image based on the updated code and dependencies.

F.5 Artifact Registry Push: The newly created container image is pushed to Google Artifact Registry, a private container registry for storing and managing container images securely.

F.6 Cloud Run Deployment: Google Cloud Run, a serverless container platform, pulls the latest container image from Artifact Registry and deploys it as a new version of the Admission Sarthi - College Predictor system.

F.7 Zero Downtime Deployment: Cloud Run automatically manages the deployment process, ensuring zero downtime by routing traffic seamlessly from the old version to the new version without interruptions for the end user.

F.8 Frontend Access: Throughout the deployment process, users continue to access the front end of the Admission Sarthi - College Predictor system without any disruptions, as Cloud Run handles traffic routing and version management transparently.

IV. METHODOLOGY

A. Data Collection and Preprocessing:

The primary data source for the Admission Sarthi - College Predictor project is the official cut-off data from the MHT-CET exam 2023, provided by the cet cell on their website. However, due to the unstructured nature of the data presented in PDF format, automated conversion into a usable format (such as CSV) was not feasible. As a result, a manual data entry process was undertaken, where the cut-off data was meticulously transcribed into Google Sheets. This manual entry ensured accuracy and completeness, as each entry was carefully validated and verified during the data entry phase.

The preprocessing phase involved thorough data cleaning to address inconsistencies, typographical errors, and missing values. Feature engineering techniques were then applied to enhance the dataset. These techniques included creating new features such as previous year cut-offs, college preferences, and student demographics. By transforming and combining the raw data into more informative features, the predictive power of the machine learning model was significantly improved. These steps were crucial in preparing a high-quality dataset for analysis and modeling.

B. Machine Learning Models: Linear Regression Approach

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In the context of Rank versus Percentile, it aims to understand how changes in Percentile influence changes in Rank, assuming a linear relationship.

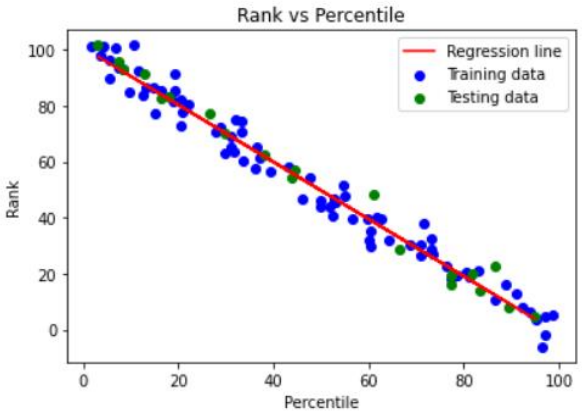

Fig 2. Rank vs Percentile Scatter Plot

The scatter plot graphically represents this relationship, with Percentile on the x-axis and Rank on the y-axis. Blue points represent training data, and green points represent testing data. The red regression line shows the predicted Rank values based on Percentile, indicating a strong inverse correlation—higher Percentile leads to lower Rank.

The model's high accuracy is evident from its training set score (R^2) of 0.976 and test set score (R^2) of 0.984. The mean squared error (MSE) of 16.34 for the test set reflects minimal prediction error, showcasing the model's reliability in predicting Rank from Percentile.

C. Web Development - Flask

The technological infrastructure supporting the Admission Sarthi - College Predictor system is built upon Flask, a lightweight and versatile Python web framework. Flask facilitated the creation of an intuitive and user-friendly interface where students can input their rank or percentile and receive personalized recommendations. Its modular design and extensive libraries allowed for efficient development and customization, ensuring the seamless integration of various features and functionalities into the portal.

D. Admission Sarthi Bot:

The Admission Sarthi - College Predictor system goes the extra mile by seamlessly integrating an OpenAI-based bot into its platform. This integration is made possible through the utilization of an OpenAI API key and Langchain-like models, which act as the backbone for AI-driven assistance and addressing student queries. By incorporating this advanced AI technology, the platform elevates its interactive nature, offering students a more dynamic and responsive user experience.

The OpenAI-based bot serves as a virtual assistant, capable of understanding natural language queries and providing intelligent responses based on deep learning algorithms. Through continuous learning and adaptation, the bot becomes increasingly proficient in addressing a wide range of student inquiries, ranging from college admission procedures to course information and eligibility criteria. This AI-driven assistance not only streamlines the user experience but also empowers students with accurate and timely information, enhancing their decision-making process.

E. Deployment:

The deployment of the Admission Sarthi - College Predictor system on Google Cloud Run with Artifact Registry is a strategic choice that brings several advantages. Google Cloud's infrastructure and services are known for their scalability, reliability, and seamless accessibility, which are crucial for a system like the Admission Sarthi - College Predictor. By utilizing Artifact Registry, the system can

efficiently manage container images, ensuring version control and streamlined deployment processes.

Cloud Run further enhances the system's capabilities by providing a platform for running containerized applications. This enables automated scaling based on demand, efficient resource utilization, and cost optimization. The robust security measures offered by Google Cloud ensure that the system remains protected against potential threats, providing a secure environment for user data.

The deployment approach on Google Cloud Run with Artifact Registry ensures optimal performance by dynamically adjusting resources and guarantees high availability, reducing downtime and ensuring continuous user access. Leveraging Google Cloud's infrastructure and services enhances user experience, improves system reliability, and simplifies resource management, contributing to the application's success.

F. CI-CD Continuous Integration & Continuous Deployment

Continuous Integration (CI) and Continuous Deployment (CD) streamline development processes by automating code integration, testing, and deployment, ensuring faster, more reliable software delivery with minimal manual intervention.

V. FLOWCHART

The following flowchart illustrates:

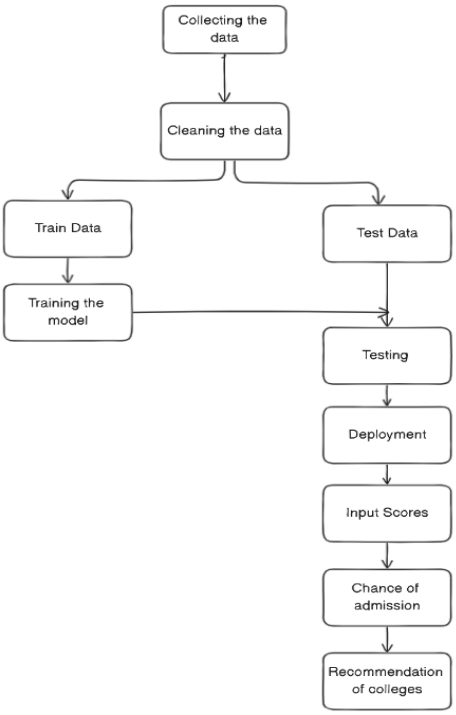

Fig 3. Machine Learning process

The process initiates with the collection of data from official sources, such as the MHT-CET exam 2023, provided by the cet cell on their website. The collected data undergoes a meticulous cleaning phase to ensure data quality and consistency. Subsequently, the dataset is divided into training and testing sets, with the testing data reserved for model evaluation, while the training data is channeled into the model training phase.

During model training, linear regression is applied to predict rank based on percentile input. Post-training, the model undergoes testing using the res testing data to ensure its efficacy in predicting admission probabilities accurately.

Upon successful testing, the application is deployed on Google Cloud Run with Artifact Registry for practical use. Users can then input their scores or percentile, triggering the generation of probability percentages for admission to various colleges. The final output includes a personalized list of recommended colleges based on the input criteria and their corresponding admission probability percentages.

CONCLUSION

In conclusion, "Admission Sarthi - College Predictor" revolutionizes the college selection process, particularly focusing on Maharashtra Engineering Colleges using MHT-CET cut-offs. By leveraging real and official data, the platform empowers students to make informed decisions regarding their college choices based on their marks or rank. It also provides probability percentages of admission into different colleges, enhancing transparency and accuracy. Additionally, the inclusion of a Generative AI-powered bot offers valuable assistance to students, making their college selection journey more efficient and less prone to errors. The project's commitment to addressing biases ensures that deserving institutions maintain their rightful place, even when faced with higher cut-offs.

Furthermore, the integration of a Mock Test feature adds another dimension to the platform, allowing students to practice and gauge their readiness for the entrance exams. This feature not only aids in preparation but also instills confidence in students as they approach their exams and college admissions. Overall, "Admission Sarthi - College Predictor" strives to empower students with comprehensive tools and insights, bridging the gap between aspirations and opportunities in higher education.

REFERENCES

[1] H. Fathiya and L. Sadath, *"University Admissions Predictor Using Logistic Regression,"* 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2021, pp. 46-51.

[2] A. Sivasangari, V. Shivani, Y. Bindhu, D. Deepa and R. Vignesh, *"Prediction Probability of Getting an Admission into a University using Machine Learning,"* 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1706-1709.

[3] M. Omaer Faruq Goni, A. Matin, T. Hasan, M. Abu Ismail Siddique, O. Jyoti and F. M. Sifnatul Hasnain, *"Graduate Admission Chance Prediction Using Deep Neural Network,"* in 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, 2020, pp. 259-262.

[4] S. Sridhar, S. Mootha, and S. Kolagati, *"A University Admission Prediction System using Stacked Ensemble Learning,"* in 2020 Advanced Computing and Communication Technologies for High-Performance Applications (ACCTHPA), Cochin, India, 2020, pp. 162-167.

[5] S. M. Tippana, S. Manju, et al., *"MBBS College Predictor,"* International Journal of Research Publication and Reviews, 2023.

[6] S. Bhoite and A. More, *"Engineering & Technology admission analysis and Prediction,"* in Gedrag & Organisatie Review, vol. 33, 2020.

[7] P. Wang and Y. Jia, *"Research on Prediction of College Students' Performance Based on Support Vector Machine,"* in Proceedings of the 2020 International Conference on Advanced Education, Management and Information Technology (AEMIT 2020), pp. 92-95, 2020.

[8] D. A. Chithra, M. Nath, P. Rohith, S. Bindushree, and S. Swaroop, *"Prediction for University Admission using Machine Learning,"* International Journal of Recent Technology and Engineering (IJRTE), vol. 8, 2020, pp. 4922-4926.

[9] Z. Wang and Y. Shi, *"Prediction of the admission lines of college entrance examination based on machine learning,"* 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 2016, pp. 332-335.

[10] A. AlGhamdi, A. Barsheed, H. AlMshjary, and H. AlGhamdi, *"A Machine Learning Approach for Graduate Admission Prediction,"* in Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing (IVSP '20), Association for Computing Machinery, New York, NY, USA, 2020, pp. 155–158,

[11] T. Lux, R. Pittman, M. Shende, and A. Shende, *"Applications of supervised learning techniques on undergraduate admissions data,"* in Proceedings of the ACM International Conference on Computing Frontiers (CF '16), Association for Computing Machinery, New York, NY, USA, 2016, pp. 412–417.

[12] M.S. Acharya, A. Armaan, and A.S. Antony, *"A Comparison of Regression Models for Prediction of Graduate Admissions,"* 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5.