JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

College Predictor using Machine Learning

Sharvari Salgaonkar[1], Akshat Patel[2], Parth Vaghela[3], Sharvari Patil[4], Sweedle Machado[5]
Student[1], Student[2], Student[3], Assistant Professor[4], Assistant Professor[5]
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

ABSTRACT
The process of choosing the appropriatecollege and course based on the obtained score and making an informed decision can be intimidating, and students sometimes struggle to estimate whichuniversities they will be admitted to. There are some existing projects that refer to this issue but they lack in encompassing all colleges present in Maharashtra and also, they do not provide the facility to choose a college which is distance friendly and saves traveling time. This research work proposes an Engineering admission recommendation system for helping students choose thebest college and branch based on their preferences for the MHT-CET examination. The system takes as input the student's percentile/rank and based upon college rankings, location, available branches, seats, and the previous year's cut-off to generate an ideal preference list and determine the chances of admission. The new dataset consisting all engineering colleges of Maharashtra has been created and Machine learningmodels such as Regression and other techniques are usedto analyse the dataset and provide suitable output. This application aims to help students secure admission to their desired college in a hassle-free manner.
Keywords: Admission recommendation system, prediction,machine learning algorithms, preference list, MHT-CET.

I. INTRODUCTION
Selecting the right college and course is a critical decision that can significantly impact a student'sacademic and professional future. However, the admission process can be overwhelming, and students often struggle with predicting the colleges that they are most likely to secure admission based on their rank andbranch preference. Consequently, students may make wrong decisions and regret their choices later. There exist plenty models that look into this issue however they consider only the top colleges in a few metropolitan cities thus limiting the implementation of the system and degrading the accuracy.
Our research work is presented to address this issue. This paper presents a college predictor tool that utilizes a machine learning algorithm to predict the colleges that a student is most likely to secure admissionbased on their MHT CET rank and branch preference. This model helps students make informed decisions by providing them with accurate information on the collegesthey can get into based on their performance. The tool takes as an input, the MHT-CET percentile/rank and useshistorical data on admission trends, cut- offs, and other related factors, and generates a list of colleges that a student can get into based on their rank and branch preference. This system takes into consideration even thesmallest towns of the state, thus extending the chance to these towns. It is trained and tested upon a vast dataset which in turn improves the accuracy of the model. This dataset includes almost all the Engineering based Colleges in Maharashtra and considering all the Minoritycategories that are applicable to these various colleges.
MHT-CET Rank is the major factor that is used to predict the desired college, as the Rank of a student is distinct whereas many students can have the same percentile, thus decreasing the accuracy of the predictions. This system predicts the preference list based on the MHT-CET Rank to provide a great accuracy of the studentso here if the student enters the percentile as input the system automatically converts percentile to Rank andoutputs the Preference List. Nowadays, many Students prefer taking admission into a college which is nearby to their residence thus avoiding the traveling inconvenience.As a complementary feature the system takes into consideration the distance between the student's residenceand the desired college thus predicting the Preference Listin an increasing order of the distances which were considered.
This paper provides an overview of the datasources, methodology, and implementation details of the tool, and presents the results of the tool's performance evaluation. Finally, we discuss the potential implications and future work in this field.

II. LITERATURE REVIEW
With the increasing demand for higher education abroad, predictive models are gaining importance in determining the chances of admission to a particular college. The research paper [1] explains a college predictor tool using machine learning techniques to assist Indian students who wish to study abroad. Previous research in this field has focused on probabilistic and classification algorithms, utilizing previous data topredict the best- suited college for a student. The system also uses K- Nearest Neighbor and Multivariate LogisticRegression algorithms to predict the likelihood of admission and the Decision Tree algorithm to suggest thelist of universities [1].
The paper [2] aims to help students fill out their Karnataka CET preference form and make better choices of college before allotment.

Previous research used technologies like Apriori, support vector machines, neuralnetworks, decision trees, logistic regression, and data mining techniques, but faced limitations such as insufficient data on students' marks and restricted access to data sets from previous years. This system employs Python machine learning libraries such as pandas and NumPy to create a database that differentiates between branches, colleges, and categories. A webpage was created where students can enter their marks, preferred branch, and college to generate a list of colleges. The authors suggest that machine learning techniques can be useful in predicting college admission by taking into account various factors such as marks, branch, and college. Overall, the paper highlights the potential ofmachine learning in automating the college admission process and improving the decision-making process for students.

The paper [3] aims to create a trusted recommender system (RS) based on data mining techniques and knowledge discovery rules to facilitate fairenrolment of students in university colleges according to standard criteria. The work proposes two cascaded hybrid recommenders: the track recommender that uses knowledge discovery rules to recommend the best track for students and the college recommender that classifies students based on various parameters such as resident status, gender, and reservation category to recommend thebest college for them. Python machine learning libraries such as NumPy, panda, pytorch, and matplotlib are used for implementing the proposed system. Overall, this worksuggests that a hybrid recommender system based on datamining techniques can improve the fairness of the college admission process by recommending suitable tracks and colleges to students. Paper [3] is based upon the predictions given by the model which in turn is trained onthe Saudi Arabia's colleges data.

In paper [4] focuses exclusively on students whowish to complete their higher education in the United States. Students then have to choose the college they wantto study at or the college they want to apply to. The application fee is high because not all universities accept applications. Here comes the problem of not knowing which university a student can get into. This article focuses on how to predict a student's college admissions likelihood based on their profile using multiple machine learning algorithms such as K nearest neighbors, linear regression, ridge regression, and random forest.

All the existing systems studied in the literature survey are either implemented for a specific state or country or designed to predict abroad universities for higher studies. Moreover, it is being observed that there isno system which focuses on the admission process of engineering admissions in Maharashtra. The current paperfocuses on almost all the colleges in Maharashtra and makes a Preference List of colleges for the students basedon their MHT-CET percentile/rank and their preferred Branch. The system is well equipped to convert percentileto rank for the ease of the student. Also, this system considers the distance of the colleges from the residence of the students thus predicting the travel efficient as well as the list of best suited colleges for the students ofMaharashtra.

III. PROPOSED ARCHITECTURE

The proposed system presents a machine learning based college predictor model designed specifically for engineering admissions of Maharashtra colleges. It provides users with a Preference List that willhelp the users to have a good chance of getting into a preferred college, based on their academic performance. Fig 1.0 shows the workflow of the proposed system.

The first and the most essential component of thesystem is the dataset. The dataset for this study is sourcedfrom the official website of the Maharashtra Common Entrance Test (MHT CET) as a PDF file. This dataset had a lot of missing data and was not as per the requirement ofthe project. The missing values were fetched and a new dataset was built by following the below steps. Firstly, thePDF file was converted to Excel format using the free online PDF editor Aspose [5], which is an open source module. The converted Excel file was processed separately, extracting required rows and merging them into a single Excel file. The row of this excel file consistedof college, branch detail. The columns of excel file had College Code, College Name, Branch, and Location, and different categories. The name of a category can divide into three parts: first part is (L for Ladies, G for General, PWD for PWD, DEF for Defense seat), second part is (OPEN, OBC, SC, ST, VJ, NT), and third part is (S for State, H for Home University, O for Other than Home University). The file was processed to maintain consistency by converting each field to string format and converting cut-off data into the

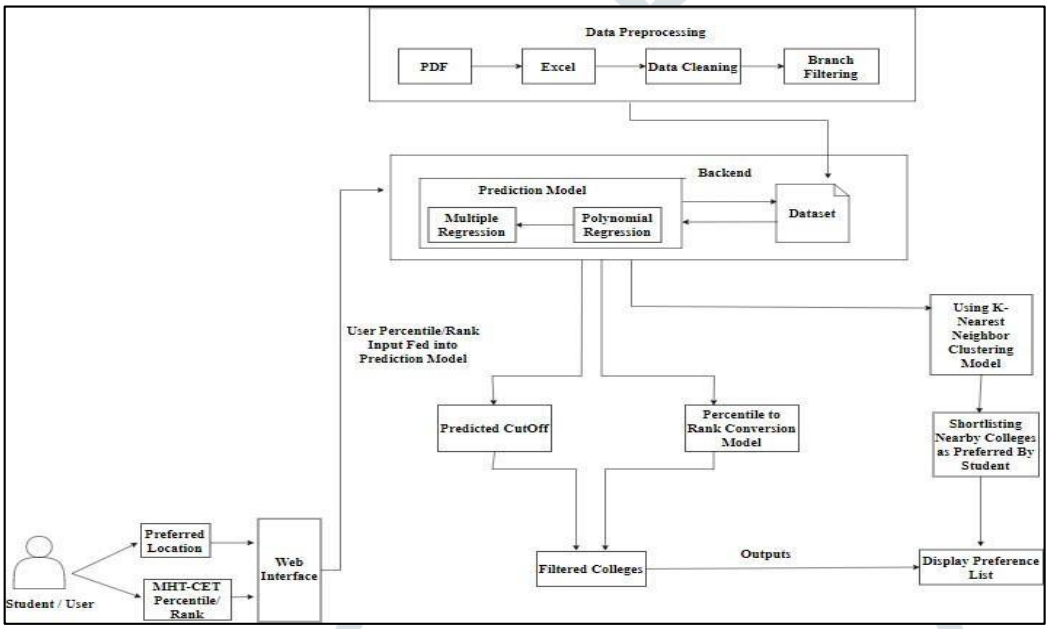

Fig 1.0 -Working of the Proposed System

same format: rank followed by percentile in a bracket. The process involved merging similar branches with different variations in name into a single branch, which was done manually for all possible branches, such as Computer Engineering and Computer Science,

to create a comprehensive list of cut-offs for the MHT-CET. This dataset was further used by the machine learning models to predict the cut-off data for the upcoming year.

Polynomial regression [6] and multiple regression [7] are preferred for prediction models due to their ability to capture complex non-linear relationships and account for multiple factors influencing the outcome (such as College branch, Rank, Percentile etc.). Other models like neural networks and decision trees are often not used due to their higher complexity and difficulty in interpreting results. Additionally, these models might require more computational resources and expertise to tune and train effectively.

The data for the year's 2019-20, 2020-21, 2021-22 was used as the training data. The cut-off was predicted for the year 2022-23 by following the trends and patternsobserved in the previous year's data. First attempts were made to plot a Percentile Vs Rank graph based on the previous year's dataset (i.e., 2019-20, 2020-21, 2021-22)as shown in Figure 1.1, Figure 1.2, Figure 1.3. The Percentile Vs Rank graph is plotted using previous years'student's CET Rank.

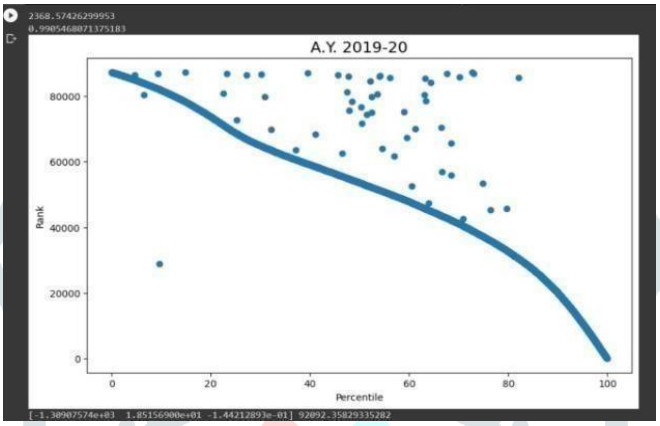

Fig 1.1- Rank Vs Percentile graph of the year 2019-20

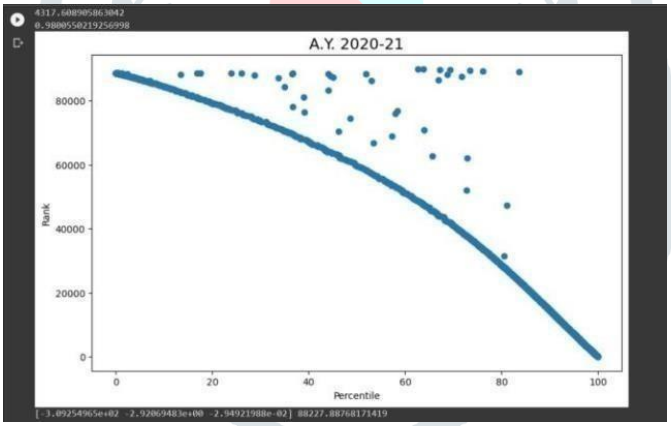

Fig 1.2- Rank Vs Percentile graph of the year 2020-21

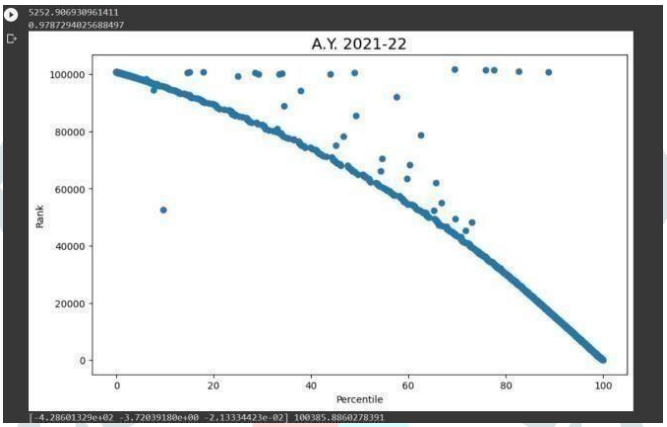

Fig 1.3- Rank Vs Percentile graph of the year 2021-22

The above Graphs are plotted using student's percentile and student's rank on X and Y axis respectively.Also, for each graph the constant value of equations is listed below in the image.

Comparisons were made between the graphs of the previous year's to further investigate the data. This comparison helped to identify any trends or patterns in thegraph, such as whether there was a general increase or decrease in the curve over time, etc. and in turn predict the graph for the upcoming year. This visualization helped to understand better the relation of percentile and rank of students over time.

Also, the graph was plotted for Rank of College Vs Rank of Student to further investigate the data and to identify any trends and

patterns which can be observed inthe year 2019-20,2020-21 and 2021-22 respectively. The observations from the plotted graphs showed that lower the student rank higher is the rank of college. For example,if a student gets the Rank 1 then the chances that the student will get the admission in the college Rank 1.

After visualizing the trends and the patterns in the previous year's graphs, polynomial regression is applied which generates a polynomial equation of degree 3 for the rank vs. percentile graph and college rank vs student rank student graph. These equations' constantvalues along with the number of students(male/female) that appeared for CET for a given year, different branches,number of categories, etc. were used as various factors for Multiple regression. The output of Multiple Regression are constant values of a polynomial equation of degree three. This equation forms the predicted graph for the upcoming year which is explained below in the predictedresults column.

This method is applied for finding the rank vs percentile graph as well as rank of college vs cut-off rankfor upcoming year i.e., 2022-23. The rank vs percentile graph is used for converting CET percentiles to rank for the next year i.e., 2022-23. By matching the rank ofcollege with the student rank in the predicted graph (college rank vs. student rank), college is decided for student.

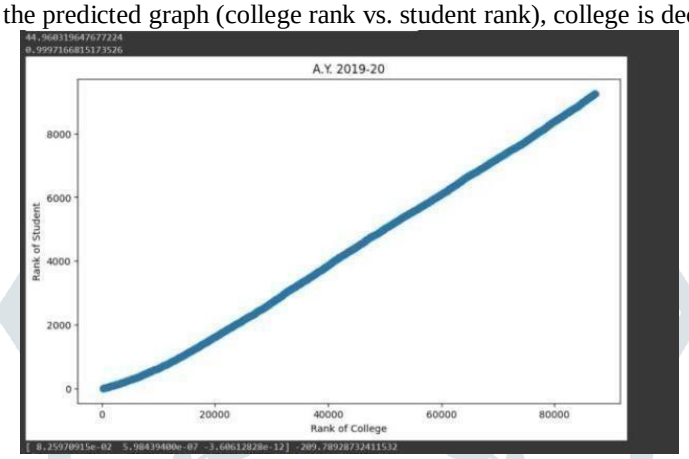

Fig 1.4- Rank of Student Vs Rank of College graph of the year2019-20

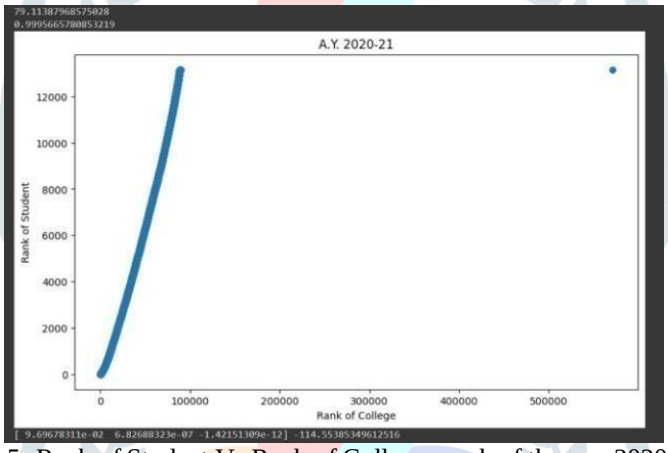

Fig 1.5- Rank of Student Vs Rank of College graph of the year2020-21

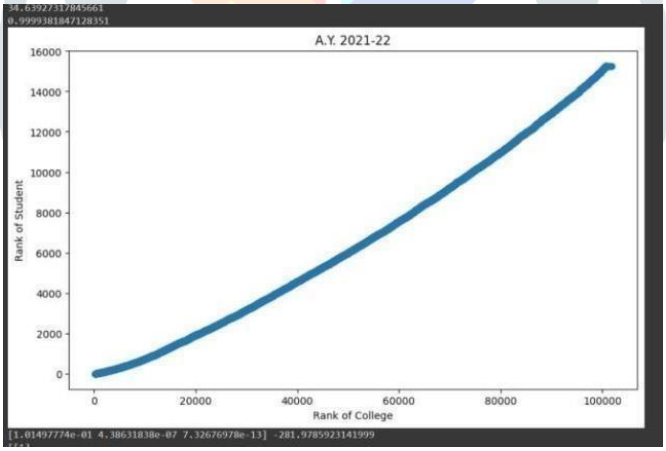

Fig 1.6- Rank of Student Vs Rank of College graph of the year2021-22

The above Graphs are plotted using college rank and student rank on X and Y axis respectively. Also, for eachgraph the constant value of equations is listed below in the image.

The front end of the system is built using React [8] a popular JavaScript library for building userinterfaces. The user interface allows users to input their academic details, such as their percentile, rank, branch, location, and category, into a form. After submitting the form, the input data is sent to the backend of the system for processing.

For backend Flask framework is used, that storesand manages the data used by the models [9]. Once the backend has processed the

user's input data, it generates alist of colleges that match the user's input details. The Preference list includes the College Code, College Name,Course, Course Code, Location, and the likelihood of theuser gaining admission to each college. This informationis then sent back to the front end of the system, where it is displayed to the user. Furthermore, based on the preferred location of the student, the list of the colleges inthe vicinity of the location provided by the student is given as output as shown in fig 2.3. This functionality makes use of K-Nearest Neighbor [10].

Overall, the architecture of the system isdesigned to be scalable, flexible, and accurate. By integrating a trained dataset with machine learning models and backend, and providing a user-friendly frontend interface, the system is able to provide users witha reliable and accurate prediction of which colleges they are likely to get into based on their academic performance.

IV. IMPLEMENTATION DETAILS

The snapshots of core modules of the proposed framework are explained below. The User Interface consists of a Landing Page which takes input from the user such as MHTCET percentile or Rank, preferredBranch, Location, Gender and the Category to which the student belongs. It displays all the Selected Branches as shown in Fig. 2.1. Processing the Input, it then outputs a Preference List of colleges that favor the student. This Preference List then can be utilized in the Official CET CELL.

Fig 2.1 - The User Interface that takes Input, the student detail.

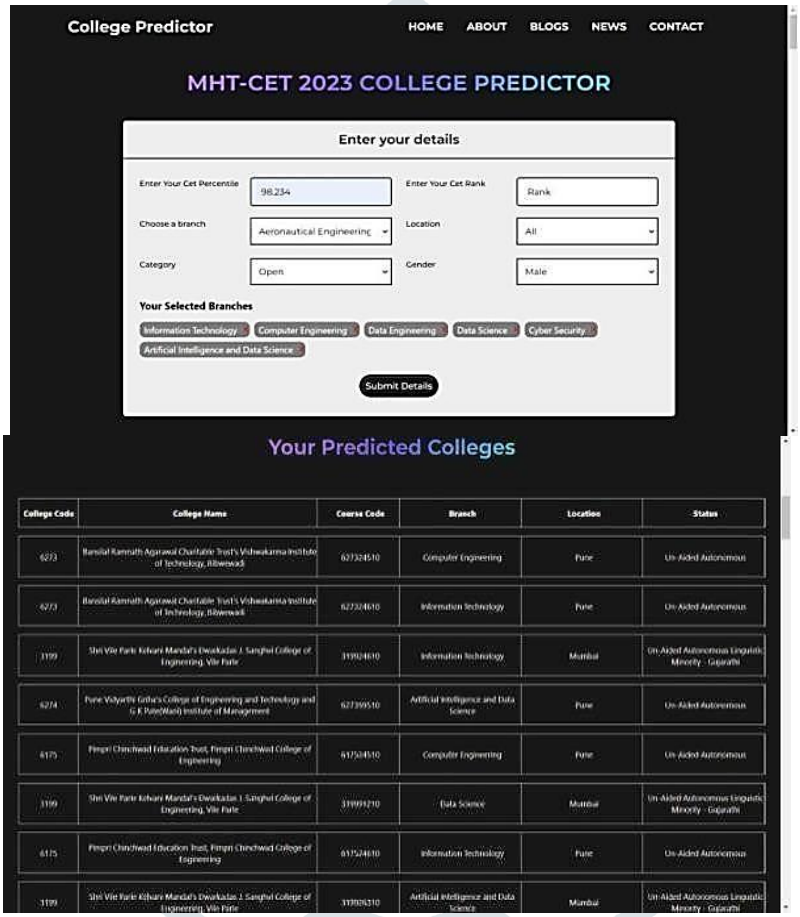

Fig 2.2 -Output given by the system in the form of Preference List which can be used in the process of Option Form Filling of Official CET-CELL.

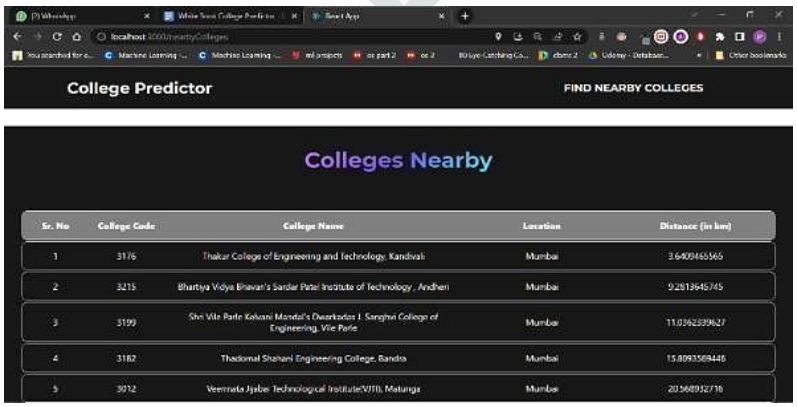

Fig 2.3 - College Preference List Predicted consideringLocation.

In Fig 2.3, Example was considered of a studentresiding in Kandivali (W), Mumbai, so nearby colleges were shown to the user.

V. PREDICTED RESULTS

In this study, the focus was on predicting the graphs for the upcoming year. Predicted graph of rank vspercentile will be used to convert student's percentile to rank. The second predicted graph of college rank vsstudent rank will be used to predicted the college that a student can get based on their CET cut-off. To make accurate predictions, first Polynomial Regression wasutilized as a statistical technique to fit a curve to the datapoints. Then multiple regression was used to predict the graph for upcoming year. If the predictions were accurate,it would increase our confidence in the models and methods used in our analysis. On the other hand, if the predictions were inaccurate, it would suggest that there were other factors influencing the data that were not accounted for such as number of students(male/female), increase/decrease in branches, etc.

The accuracy of the predictions was evaluated by comparing the predicted values to the actual values inthe data set. Based on the analysis, the Polynomial Regression model used was able to accurately predict therelationship between Rank and Percentile, with an accuracy rate of up to 85%.

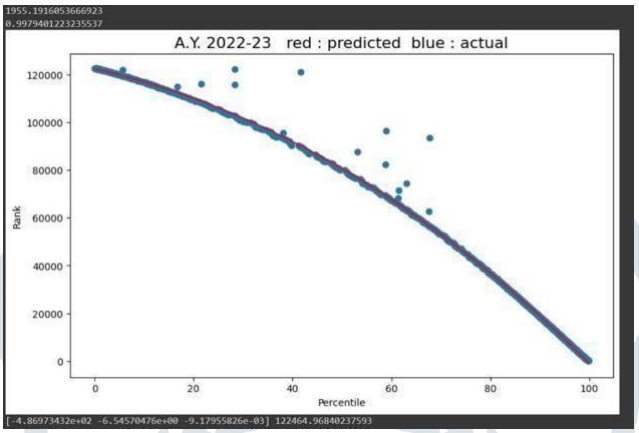

Fig 3.1

```
R sqaured: 85.11
Mean Absolute Error: 3.0489239348743
Mean Square Error: 2.636034234382
Root Mean Square Error: 2.358304930293
```

Fig 3.2

Fig 3.1 is the predicted graph of student's rank vs. percentile for the upcoming year 2022-23 and Fig 3.2 shows the accuracyof the fit.

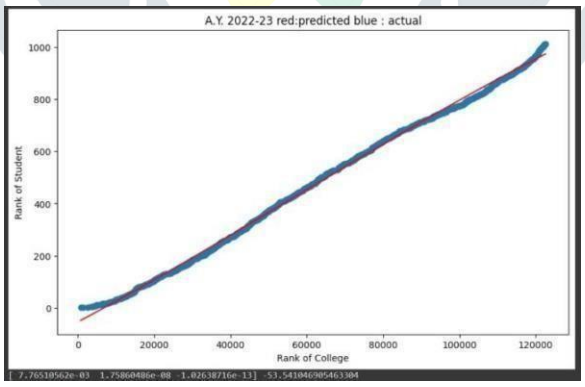

Fig 3.3

```
R sqaured: 92.34
Mean Absolute Error: 3.0489239348743
Mean Square Error: 2.636034234382
Root Mean Square Error: 2.358304930293
```

Fig 3.4

Fig 3.3 is the predicted graph of Rank of college Vs rank of Student for the upcoming year 2022-23 and Fig 3.4 shows theaccuracy of the fit.

VI. CONCLUSION

The main objective of this research was to develop a system that can be used by students for gettingtheir admission to Undergraduate College, based upon their MHT-CET Rank. Polynomial Regression was used for evaluating their chances of getting into a college withan average accuracy of 85%. A simple user interface wascreated with the help of React as Frontend and Flask for backend. The overall system was successful as students can save time and can make better and fast decisions. This system can be further implemented on an India- level exam i.e., JEE MAINS by enlarging thedataset by including the colleges in India, thus widening the scope of the system.

REFERENCES

[1] Sonawane, H. (2017). *Student admissionpredictor* (Doctoral dissertation, Dublin, National College of Ireland).
[2] HK, V. S., Neeraj, S., & Shashidhar, R. (2021). An Automated Prediction Model for College Admission System. *Elementary Education Online*, *20*(6), 1172- 1172.
[3] Ragab, A. H. M., Mashat, A. F. S., & Khedra, A. M. (2014). Design and implementation of a hybrid recommender system for predicting college admission. *International Journal of Computer Information Systems and Industrial ManagementApplications*, *6*, 35-44.
[4] Apoorva, D. C., Chandu Nath, M., Rohith, P., Shree, S. B., & Swaroop, S. (2020). Prediction for university admission using machine learning. *Int. J. Recent Technol.Eng*, *8*(6).
[5] Aspose: https://www.aspose.com.
[6] Polynomial Regression: https://www.javatpoint.com/machine-learning polynomial-regression
[7] Multiple Regression: - https://byjus.com/maths/multiple- regression/
[8] React: https://react.dev/
[9] Flask: https://flask.palletsprojects.com/en/2.3.x/
[10] K-Nearest Neighbor:-https://www.javatpoint.com/knearest-neighbor algorithm for-machine-learning
[11] Wang, Z., & Shi, Y. (2016, October). Prediction of the admission lines of college entrance examination basedon machine learning. In *2016 2nd IEEE InternationalConference on Computer and Communications(ICCC)* (pp. 332-335). IEEE.
[12] Mirji, A., Kakade, M., Paithankar, D., & Shirbhate,V. COLLEGE PREDICTOR AND ANALYZER.
[13] Kireeti, G. S. K., Prithvi, J., Divya, M., & Kumari, C.U. (2023, April). Predicting Employability and Admissionfor MS Students using ML Regression Models. In *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)* (pp. 1-8). IEEE.
[14] Sivasangari, A., Shivani, V., Bindhu, Y., Deepa, D.,& Vignesh, R. (2021, April). Prediction Probability of Getting an Admission into a University using Machine Learning. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1706-1709). IEEE.
[15] Kumari, K., Kataria, M., Limbani, V., & Soni, R. (2019, April). CAPSLG: College Admission Predictor and Smart List Generator. In *2nd International Conference on Advances in Science & Technology (ICAST)*.
[16] Cs, K. (2014). University Admission Prediction usingMachine Learning. *Global Journal of Research and Review*.
[17] Sujay, S. (2020). Supervised machine learning modelling & analysis for graduate admission prediction. *Int. J. Trend Res. Dev*, *7*(4), 5-7.
[18] Wang, Z., & Shi, Y. (2016, October). Prediction of the admission lines of college entrance examination based on machine learning. In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)* (pp. 332-335). IEEE.
[19] Zhang, Y., Feng, X., Qu, X., Wang, S., Sun, L., Hua, P., & Wang, Y. (2022, January). Research on thePrediction Method of the College Professional Admission Scores. In *2022 International Seminar on Computer Science and Engineering Technology (SCSET)* (pp. 406-409). IEEE.
[20] AlGhamdi, A., Barsheed, A., AlMshjary, H., &AlGhamdi, H. (2020, March). A machine learning approach for graduate admission prediction. In *Proceedings of the 2020 2nd International Conferenceon Image, Video and Signal Processing* (pp. 155-158).