

IIT and NIT College Prediction System

Abstract:

In the present conditions, students regularly have difficulty finding a fitting institution to pursue higher studies based on their profile. There are some advisory administrations and online apps that recommend universities but they ask huge consultancy fees and online apps are not accurate. So, the aim of this research is to develop a model that predict the percentage of chances into the university accurately. This model provides also the analysis of scores versus chance of prediction based on historical data so that students can understand whether their profile is suitable or not. The proposed model uses linear regression and random forest algorithms but cat boost algorithm is giving highest accuracy.

The college admissions process can be daunting for many students, and even though some organisations offer admission assistance, the advice they provide is not always accurate and often expensive. This study aims to develop a model that can more accurately predict a student's chances of college admission.

I.INTRODUCTION

Anyone pursuing their postgraduate studies, it would be difficult for them to find out what college. People may apply to many universities that look for candidates with a higher score set, instead of applying to universities at which they have a chance of getting in to. This would be detrimental to their future. It is very important that a candidate should apply to colleges that he/she has a good chance of getting into, instead of applying to colleges that they may never get into. The Education Based Prediction System helps a person decide what colleges they can apply to with their scores.

The dataset that is used for processing consists of the following parameters: University name, age, percentage, category, gender. The GRE Test (Graduate Record Examinations) is a standardized test used by many universities and graduate schools around the world as part of the graduate admissions process. The dataset that is used for processing consists of the following parameters: University name, age, percentage, category, gender. The GRE Test (Graduate Record Examinations) is a standardized test used by many universities and graduate schools around the world as part of the graduate admissions process.

II. MOTIVATION

- **High Competition:** Admission to IITs and NITs in India is extremely competitive, with millions of students using for a limited number of seats. A prediction system can help students understand their chances of admission, reducing anxiety and stress.
- **Informed Decision-Making:** Prospective students can make more informed decisions about their choice of courses and colleges. They can assess the probability of getting into their preferred institution, adding them in choosing the right path for their academic and career goals.
- **Time and Resource Efficiency:** Applying to multiple colleges is time-consuming and expensive. A prediction system can help students focus their efforts on institutions where they have a higher chance of admission, saving both time and money.

III. LITERATURE REVIEW

Prediction Probability of Getting an Admission into a University using Machine Learning: A machine learning model for estimating the odds of admission to a university is constructed as part of the project. In an 80:20 ratio, the dataset was split into a training set containing 400 profiles and a testing set containing 100 profiles. The authors underlined the significance of pre-processing in cleaning and preparing the data for the prediction algorithm. They observed that the data acquired from Occidental College required refinement to be appropriate for machine learning techniques. The authors recognized the issue of missing data and used a variable form to find the missing data and treat it. The data were then pre-processed and randomly divided into training and testing sets, with the training collection using 80.[1] [2019]

Prediction Probability of Getting an Admission into a University using Machine Learning: Accurate prediction of the results of the college entrance examination (CEE) is crucial for candidates to complete applications and conduct pertinent analyses of the CEE. Although data analytics, probability models, and a few weighted combination models are used to predict CEE scores, the reference value is minimal, and the accuracy of current prediction methods is relatively low. In this study, the lines of research and prediction for college admission are carried out using machine learning techniques, particularly the Adaboost method, which is a part of ensemble learning. This model is shown to provide more accurate predictions than current methodologies, demonstrating its potential for improving college admissions processes.[2] [2021]

University Admissions Predictor Using Logistic Regression: The approach of machine learning (ML), which enables computers to learn and evolve on their own utilizing data, is discussed in the paper. ML may be used for many things, including identifying patterns and trends, automating decision-making, and reaching large-scale data-driven outcomes. ML models can be classified into three categories: reinforcement learning, unsupervised learning, and supervised learning. However, ML has drawbacks, such as the need for correct data and the time-consuming process of determining the optimum algorithm for a given task. The steps involved in generating an ML model are data retrieval, pre-processing, and analysis, model creation, deployment, testing, and iteration. Notwithstanding these difficulties, ML has created opportunities for tackling issues when the answer is obvious but the way to accomplish it is still not.[3] [2021]

University Admissions Predictor Using Logistic Regression: Proposed a University Admission Predictor (UAP) system. They have tried to determine the predictions using various machine learning algorithms such as KNN, Decision Trees, Ridge Regression etc to predict the admission chances using the variables such as GRE, TOEFL, GPA, SOP, LOR etc. During the research they have tried to compare the algorithms mentioned above for choosing the best for their proposed UAP system. The proposed system is able to predict the admission chances with 79% of average accuracy.[4] [2020]

The proposed system used the means of database. It transformed there optional database for all the specific and relevant data of students into a universal database Then the proposed system with the help of an algorithmic model calculatessimilarities between the data of students who got successful admissions and the student whowants to get the admission based on the data provided by the willing student The proposed system used KNN or K-nearest neighbour machine learning algorithm for the similarities between the data of students who got successful admissions and the student who wants to getthe admission based on the data provided by the willing student The proposed system used KNN or K-nearest neighbour machine learning algorithm for the calculation of top K universities for the N similar users.[6] [2018]

IV.METHODOLOGY

The "College predictor and analyzer" project aims to predict which colleges a student is likely to get admitted to, based on their academic credentials and other relevant factors. Here is a proposed methodology for the project:

Algorithms

a. Decision trees(2020)

A predictive model was developed using decision trees to predict the probability of admission to a college based on several factors such as the student's high school grades, standardized test scores, extracurricular activities, and demographic information. The results showed that the predictive model was accurate and could be used as a tool for college admission.-It provides 80% accuracy.

b. K-Nearest Neighbours

It is an algorithm which is used widely for classification and regression problems. Due to its simplicity and effectiveness, it is easy to implement and understand. It is a supervised machine learning algorithm that uses available data to create the model and further that model can be applied to classify the new data. The class of new data is determined by the class of its neighbours. Distance is calculated between the unseen data sample and the all other data samples already present in the data-set.- It provides 90% accuracy.

V.Models

Module1: Exploratory Data Analysis

Module 2: Pre – Processing

Module 3: Data Visualization

Module 4: Training and Testing Dataset

Module 5: Creating Model

Module 6: Optimizing the model

Module 7: check accuracy

Module1: Exploratory Data Analysis

Exploratory Data Analysis, or EDA, is a crucial step in the data analysis process that involves studying, exploring, and visualizing information to derive important insights. To find patterns, trends, and relationships in the data, it makes use of statistical tools and visualizations. This helps to formulate hypotheses and direct additional investigations.

Module 2: Pre – Processing

Pre-processing in Python for a college prediction system for NIT (National Institutes of Technology) and IIT (Indian Institutes of Technology) involves cleaning, transforming, and preparing the data before feeding it into a machine learning model.

Unnamed: 0	year	institute	type	round no	quota	pool	institute short	program name	program duration	degree short	category	opening_rank
0	0	2016	IIT	6	AI	Gender-Neutral	IIT-Bombay	Aerospace Engineering	4 Years	B.Tech	GEN	838
1	1	2016	IIT	6	AI	Gender-Neutral	IIT-Bombay	Aerospace Engineering	4 Years	B.Tech	OBC-NCL	408
2	2	2016	IIT	6	AI	Gender-Neutral	IIT-Bombay	Aerospace Engineering	4 Years	B.Tech	SC	297
3	3	2016	IIT	6	AI	Gender-Neutral	IIT-Bombay	Aerospace Engineering	4 Years	B.Tech	ST	79
4	4	2016	IIT	6	AI	Gender-Neutral	IIT-Bombay	Aerospace Engineering	4 Years	B.Tech	GEN-PWD	94

Module 3: Data Visualization

Visualizing data in Python using Matplotlib, Pandas and Seaborn. Data visualization is the discipline of trying to understand data by placing it in a visual context so that patterns, trends, and correlations that might not otherwise be detected can be exposed. Data visualization plays a significant role in the representation of both small and large data sets, but it is especially useful when we have large data sets, in which it is impossible to see all of our data.

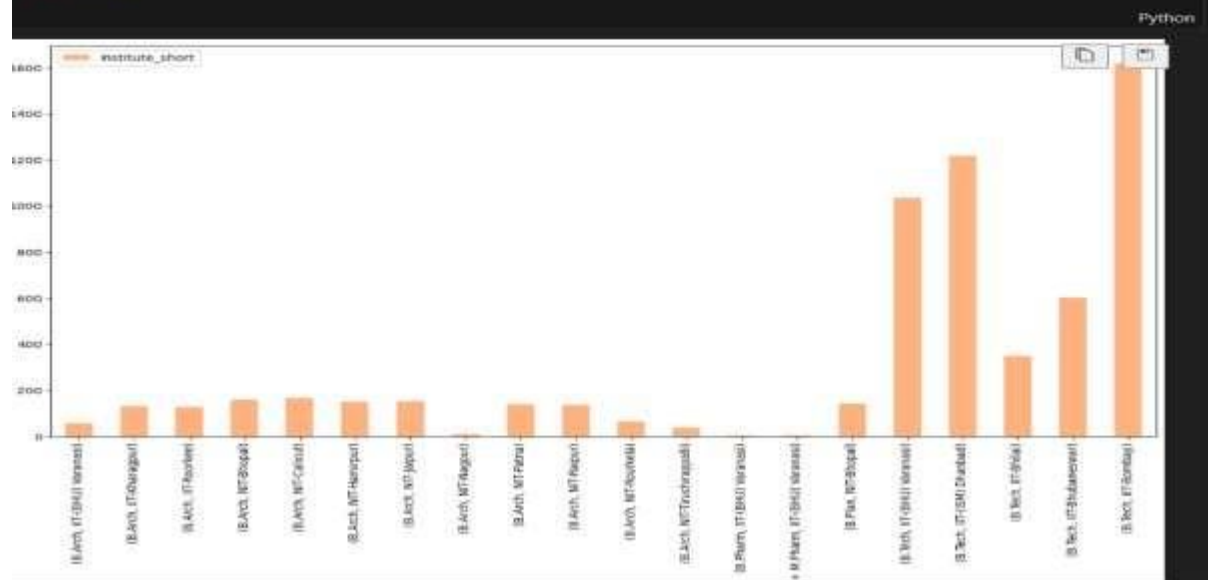
```
# it returns the summary statistics of numerical column
df.describe()
```

	year	round_no	opening_rank	closing_rank	is_preparatory
count	64958.000000	64958.000000	6.495800e+04	6.495800e+04	64958.000000
mean	2020.421580	2.609348	8.259642e+03	1.070497e+04	0.047631
std	1.149762	2.422558	2.679448e+04	3.788101e+04	0.212985
min	2016.000000	1.000000	0.000000e+00	0.000000e+00	0.000000
25%	2020.000000	1.000000	6.710000e+02	8.320000e+02	0.000000
50%	2021.000000	1.000000	2.309000e+03	2.764500e+03	0.000000
75%	2021.000000	6.000000	6.932000e+03	8.190000e+03	0.000000
max	2021.000000	7.000000	1.082601e+06	1.144790e+06	1.000000

Module 4: Training and Testing Dataset

Creating a training and testing dataset for predicting college admission into IITs (Indian Institutes of Technology) and NITs (National Institutes of Technology) would require certain data points to be collected. Here's a simplified example of how you might structure such a dataset:

```
plt.figure(figsize=(15,8))
year_club = df.groupby(['degree_short','Institute_short']).Institute_short.count().head(20).plot(kind='bar')
plt.legend()
plt.show()
```



Module 5: Creating Model

1. **Load Data:** Load your dataset containing student information and their acceptance into NIT and IIT.
2. **Split Data:** Split the dataset into features (X) and the target variable (y).
 - a. X contains student attributes (e.g., academic scores, extracurricular activities).
 - b. y contains the target variable indicating acceptance into NIT or IIT.

Module 6: Optimizing the model

Model optimization is a critical aspect of ML(machine learning) that focuses on improving the performance and accuracy of predictive models. By fine-tuning model parameters, exploring optimization techniques, and evaluating different strategies, we can enhance our models and unlock their full potential.

```
#imports

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

Module 7: check accuracy

To check the accuracy of your college prediction system for NIT (National Institute of Technology) and IIT (Indian Institutes of Technology) in Python.

- **Prepare Data:** Ensure you have a dataset with features such as scores, ranks, or any other relevant information about students and their corresponding colleges (NIT or IIT).
- **Train Model:** Use machine learning algorithms (such as logistic regression, decision trees, random forests, etc.) to train a model based on your training data.
- **Evaluate Model:** Once your model is trained, use the testing set to evaluate its accuracy. You can calculate metrics like accuracy, precision, recall, F1-score, etc.



The screenshot shows a web application titled "College Predictor System Software". It features a navigation bar with links for Home, Topics, About, and Contact Us. The main form is a blue-bordered box containing several input fields and dropdown menus. Below the form is a yellow button labeled "Tap To Predict College". The results of the prediction are displayed below the button, showing the predicted college, degree, and course. At the bottom, there are "Older" and "Newer" buttons for navigating between predictions.

College Predictor Home Topics ▾ About Contact Us

College Predictor System Software

Type your name Choose your Institute ▾

Your school name Preferred Round Number ▾

Select your Category ▾ opening rank

Select your Quota ▾ Closing rank

Choose Your Gender ▾ Would you like our website

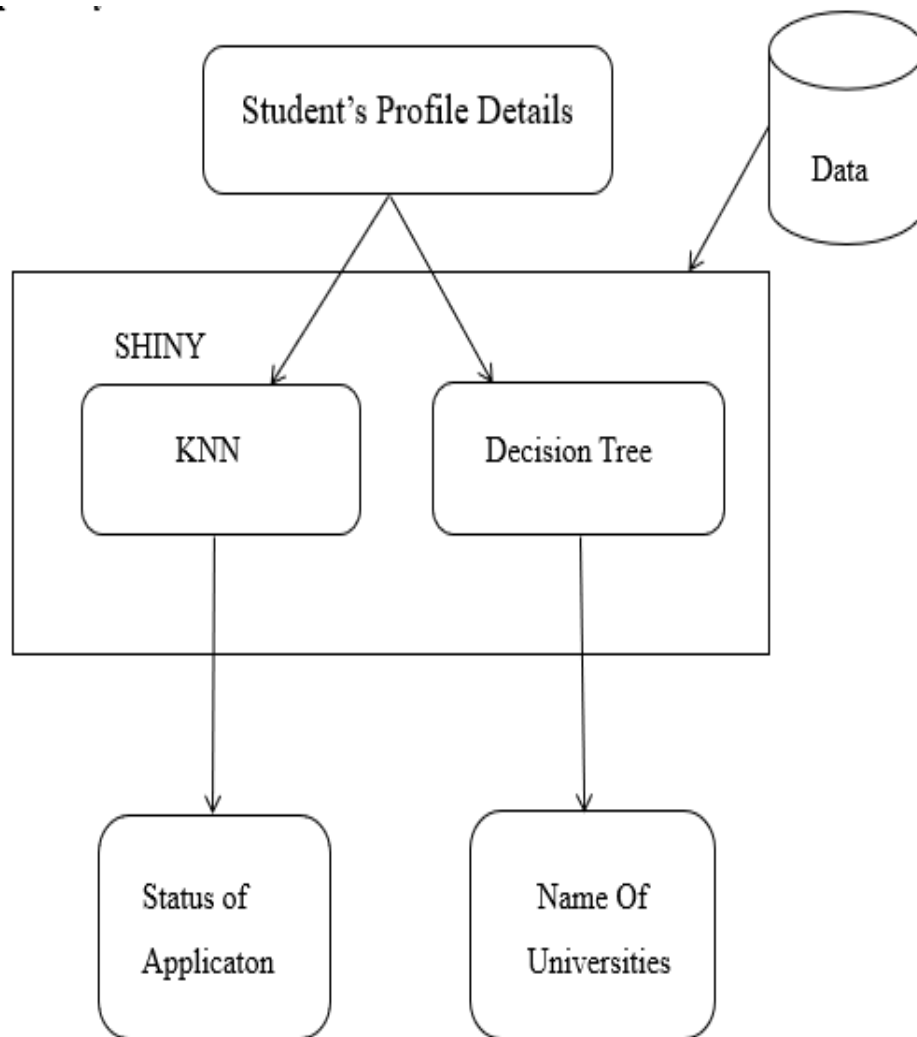
Tap To Predict College

College : IIT-(BHU) Varanasi , Degree : B.Arch , Course : Architecture

Thank you, Hope this will match your requirement !!!

Older Newer

VI. ARCHITECTURE



VII. SCOPE

- It helps student for making decision for choosing a right college.
- It Saves Students extra amount of time and money that they would spend on education consultants and application fees for the universities where they have fewer chances of securing admission.

VIII. CONCLUSION

The purpose of the project is to develop a platform to predict ranks and predicting college based on that ranks considering various parameters. In first phase, we have worked upon the scrapping of data from various websites required for the database, and created structed mongo database. We have created design using Python. We have added functionality of searching college. We are learning various ML algorithms and will finalize the algorithm based on the accuracy.

The purpose of the project is to develop a platform to predict ranks and predicting college based on those ranks considering various parameters an with increasing number of international students choosing the United States as their preferred study destination, the admissions process for these students has become more challenging. This study aims to develop a system that utilizes machine learning techniques, referred to as the Student Admission Predictor.

IX. REFERENCES

1. <https://ieeexplore.ieee.org/document/9418279>
2. <https://www.semanticscholar.org/paper/Prediction-Probability-of-Getting-an-Admission-into-Sivasangari-Shivani/02d75e4d4bf83cafbf674c2121bcfceeaa->
3. 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, doi: 10.1109/ICCIKE51210.2021.9410717
4. Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S, "Prediction for University Admission using Machine Learning" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-6 March 2020
5. H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support DecisionMaking in University Admission Systems," in IEEE Access, vol. 8, pp. 55462-55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
6. Annam Mallikharjuna Roa , Nagineni Dharani , A. Satya Raghava , J. Buvanambigai , K. Sathish,"College Admission Predictor" Journal of Network Communications and Emerging Technologies (JNCET) www.jncet.org_Volume www.jncet.org Volume 8, Issue 4, April (2018)