Name of the paper: Prediction for University Admission using Machine Learning

Year of Publication: 2020

Author: Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S

Journal Name: International Journal of Recent Technology and Engineering (IJRTE)

Theme:

Many colleges in the U.S. follow similar requirements for student admission. Colleges take different factors into account, such as the ranking on aptitude assessment and academic record review[3]. The command over the English language is calculated on the basis of their performance in the English skills test, such as TOEFL and IELTS. The admission committee of universities takes the decision to approve or reject a specific candidate on the basis of the overall profile of the applicant application.

Inference:

Numerous programs and studies have been carried out on topics relating to university admission used many machine learning models which helps the students in the admission process to their desired universities. Previous research done in this area used Naive Bayes algorithm which will evaluate the success probability of student application into a respective university but the main drawback is they didn't consider all the factors which will contribute in the student admission process like TOEFL/IELTS, SOP, LOR and under graduate score. Bayesian Networks Algorithm have been used to create a decision support network for evaluating the application submitted by foreign students of the university.

This model was developed to forecast the progress of prospective students by comparing the score of students currently studying at university. The model thus predicted whether the aspiring student

should be admitted to university on the basis of various scores of students. Since the comparisions are made only with students who got admission into the universities but not with students who got their admission rejected so this method will not be that much accurate.

Problem Understanding: Initially first we have to spend some time on what are the problems or concerns students having during their pre admission period and we should set the solutions to those problems as objectives of this research.

Data Understanding: Data should be collected from multiple sources like yocket and also consider all the factors including which will play a tiny role in student admission process.

Data Preparation: Data should be cleaned that is removing the noise in the data and filling the missing values or extreme values and finalising the attributes/factors which will have crucial importance in student admission process.

Building Models: Several ML models have to be developed using various machine learning algorithms for admission to a particular university and the user interface has to be developed to access those models.

Evaluation: Developed models are evaluated according to their accuracy scores. Once the model is finalised that model will be merged with node red for final deployment.

Data Cleaning and Analysis:

• Inspecting feature values that help identify what needs to be done to clean or pre-process until you see the range or distribution of values typical of each attribute.

• You may find missing or noisy data, or anomalies such as the incorrect data form used for a column, incorrect measuring units for a particular column, or that there are not enough examples of a specific class.

• You can know that without machine learning, the problem is actually solvable.

The data cleaning process has several key benefits to it:

1. This eliminates major errors and inconsistencies which are unavoidable when dragging multiple data sources into one dataset.

2. Having data cleaning software will make everyone more effective as they will be able to get easily from the data what they need.

3. Fewer mistakes mean happy clients, and less unhappy workers.

4. The ability to chart the various functions, and what your data is supposed to do and where it comes from your data.

• There are no missing values and outliers because we analysed the data, so for this data there is no need to fill the missing values and deal with outliers. If there are any missing values and outliers we can fill (or) drop using the fillna method and drop method and we can also standardize the data using the min-max scaler, if necessary.

Data Visualization:

• After analysing the data, we will be able to know what the features and labels are, so from the above data, the label we have to consider is Chance of Admission and then we have to consider the parameters that influence or play a major role in Chance of Admission

• We can get to know certain features that are more affected by the visualization (or) analysis or the use of feature importance method in decision tree