Machine Learning Engineer Nanodegree

Capstone Proposal

Pavankalyan kanakam

Feb,2019

Proposal:Admission prediction

Domain background

To apply for a master's degree is a very expensive and intensive work. With this idea, students will guess their capacities and they will decide whether to apply for a master's degree or not.

This is probably a question that every aspiring MS aspirant wants to know. Is my profile good enough to get a good college? Being an aspirant myself even I also have so many doubts, whether my CGPA is good enough, how should I write a solid SOP, etc.

The main goal of this problem is to predict the 'Chance of Admit' of a student in a particular university given various parameters.

Problem Statement

- The main aim of this problem is to predict the 'Chance of Admit' with high accuracy by applying various ML Algorithms and then comparing their scores..
- Compare different models to check for best model depending on r_squared score

Datasets and Inputs

This dataset is created for prediction of graduate admissions and the dataset link is below:

- https://www.kaggle.com/mohansacharya/graduate-admissions

Features in the dataset:

- GRE Scores (290 to 340)
- TOEFL Scores (92 to 120)
- University Rating (1 to 5)
- Statement of Purpose (1 to 5)
- Letter of Recommendation Strength (1 to 5)
- Undergraduate CGPA (6.8 to 9.92)
- Research Experience (0 or 1)
- Chance of Admit (0.34 to 0.97)

Additional information

- There are 9 columns: Serial No., GRE Score, TOEFL Score, University Rating, SOP, LOR , CGPA, Research, Chance of Admit
- There are no null records. It's good.
- There are 400 samples in total. That's enough.

Solution Statement

The main aim of this problem is to predict the 'Chance of Admit' with high accuracy by applying various ML Algorithms and then comparing their scores..

In this problem we use both classification and regression models

Algorithms Considered:

- Linear Regression
- Decision Tree Regressor

Benchmark Model

The benchmark model I will choose is Deciision tree regressor which gives r-squared score of 65.81%

In order to beat this I will choose anthor regression model called linear regression

While iam using linear regression I successfully got 82.14% which is far better than the decision tree regressor.

Evaluation Metrics

In this I will compare different models to check for best model depending on r-squared score.

Project Design

Step-1:loading and cleaning the data

First, I will load the data and check for whether dataset need cleaning or not.

Step-2: finding the Parameters with High Correlation against 'Chance of Admit'

Here I will Check for any Linear Relationship between the given Parameters

- By constructing Pairplot
- By constructing Correlation heatmap

Later I will find the parameters which effects the final value of chance of admit .

Step-3:applying ML algorithms :

Next, I will apply the above mention algorithms on dataset and I will comapre the accuracy of those algorithms in order to find the best one.