Machine Learning on Admission Prediction

1.Kishor R, 2.Mahesha K, 3.Ganesh Acharya,
Co author. MAHENDRA KUMAR B,

1,2,3, PG SCHOLORS, DEPT. OF MCA, DSCE
CA -ASST. PROF., DEPT. OF MCA,DSCE

I. KEY WORDS

DATA SCIENCE, LOGISTIC REGRESSION, MACHINE LEARNING, BUSINESS INTELLIGENCE, ANALYSIS.

II. ABSTRACT

This mini project developed under a group of team with the guidance of experts which predicts the admission of the given set of students are likely to be admitted or be considered as rejected. This is not built to take decision whether or not the student get placed or not but help the department to make a clear decision. This project describes admission prediction, a statistical machine learning system developed to support the work of the graduate admissions committee at the University in -ref[5] selecting the deserved students . In recent years, the number of applications to the Masters and PhD program has become too large to manage with a traditional review process. This project uses historical admissions data to predict how likely the committee is to admit each new applicant. -ref[4]

III. INTRODUCTION

This project describes admission prediction, a statistical machine learning system developed to support the work of the graduate admissions committee at the University in -ref[5] selecting the deserved students . In recent years, the number of applications to the Masters and PhD program has become too large to manage with a traditional review process. This project uses historical admissions data to predict how likely the committee is to admit each new applicant. It reports each prediction as a score similar to those used by human reviewers, and accompanies each by an explanation of what applicant features most influenced its prediction. This project makes the review process more efficient by enabling reviewers to spend most of their time on applicants near the decision boundary and by focusing their attention on parts of each applicant's file that matter the most. -ref[4]

IV. WORKING CRITERIA

As a machine learning problem, this project frames graduate admissions as probabilistic binary classification. For training data, the system reads an internal departmental database of past admissions decisions, which contains an atomized version of each applicant's file and a binary label indicating whether or not the person was admitted to the graduate program. Each student's file is represented as a high- dimensional feature vector that encodes the institutions previously attended, GPA's, test scores, letters of recommendation, area of research interest, and preferred faculty adviser. Given the historical data, the goal is to predict the probability that the admissions committee will admit each new applicant to the program.

Internally, it is implemented with an regularized logistic regression model. The regularization system acts as the feature selection mechanism in practice, producing a sparse model that uses only a small subset of highly predictive features. Upon inspection, the model focuses on much of same information that human committee members use when reviewing applications. In particular, this project prefers applicants with high GPAs and test scores, backgrounds from reputable institutions, and recommendation letters that support the applicant's potential as a researcher. -ref[3]

V. FIGURES AND TABLES

V.I FIGURES

1. DATA SCIENCE LIFE CYCLE

Every project must follow the flow of the development which is domain as Life cycle to yield the accurate outcomes as predicted. This life cycle defines the start and end of the project with step by step procedures to reach the desired outcome.

Life cycle of any project in machine learning defines the start to the end process of the flow of prediction . This means the life cycle defines from the state of understanding the data and its problem , collecting the data from the source which can be from the client side or from the servers, preparing the collected data from source which exist in different format and dimensions, splitting the data into train and tests data-sets which helps in comparing the data after applying the statistics to the trained data and to the test data, selecting the desired model and its understandings, train that selected model with the data for the machine, evaluating the trained data with the test data, further en-chancing the data for better predictions if required , and finally the deployment of that model into practice.

This life cycle makes the predictions more accurate and more efficient in training the machine with the data. It also makes the

developer more understandable and clarity in the model being developed.

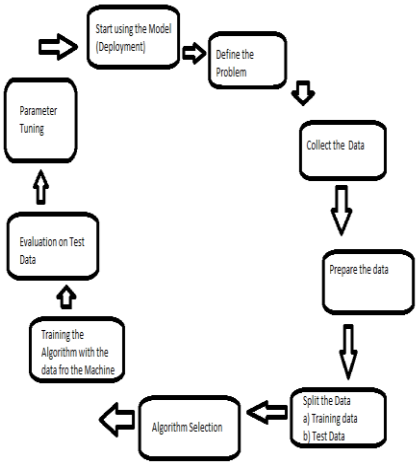

Fig.1. Data Science life cycle (Arrow marks shows the direction of flow of the process)

2. DATA QUALITY

It is very important to know the quality of the data and what percent of analysis is required to be implement on that data is also important. There are certain measures that shows the attributes of the quality aspects in form of diagrams and graphs, we use the mat-plot libraries to plot all the graphs to keep the quality in concern.

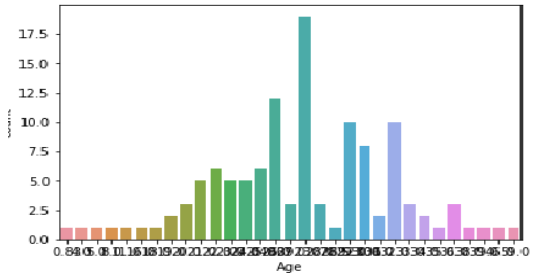

Fig 2. The dist-plot graph (shows the quality of data of different columns in the table.)

3.Distribution of data

```
<matplotlib.axes._subplots.AxesSubplot at 0x22e16247dd8>
```

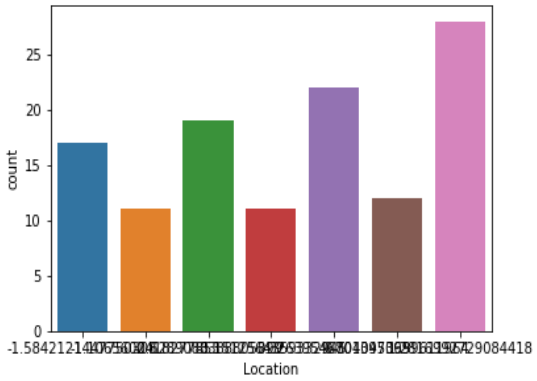

Fig 3- (Shows the various data distribution and its frequency count on different location)

The above graph the representation of various location and their distribution frequency. On x-axis is the location and the x-axis is the corresponding count frequency.

The graph at the location '9084418' is high in the frequency count and other locations are also showing the various distribution regardless the different aspects.

4. After processing of the data distribution

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x22e15af94e0
```

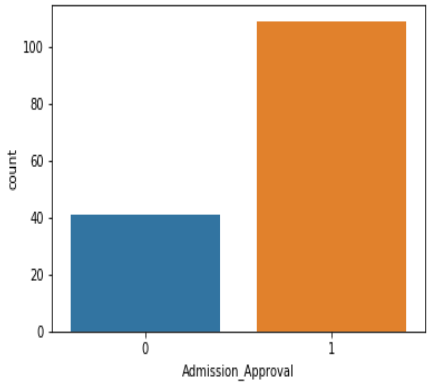

Fig 4- (Shows the data distribution on the splitting data)

V.II TABLES

Table I: DATA-SETS(sample data-sets as an example).

Applicant_Name	Applicant_ID	Percentage	Location
Kishor	120	21	Bangalore
Mahesha	121	22	Bangalore
Ganesh	122	21	Bangaore
Ravish Babu	123	22	Bangalore
Isaq Khan	124	21	Bangalore
Manjunath	124	30	Bangalore
De villiers	125	23	South Africa
Paul Walker	126	37	America
Vin Diesel	127	38	America

VI. MODEL DEVELOPMENT

Machine learning classification technique is a supervised learning that is designed to infer class labels from a well-trained set of data having dependent and independent data.

After cleaning the data as the required selection model and its analysis.

The model is selected based on the classification of data and its linearity to approach the accurate score predictions.

In this machine learning the selected model is the linear regression model based on the distribution of data levels. It is a supervised machine learning model which is used when the independent data(inputs) and the dependent data(target data) is known or given.

Since the given data-set is based on the distributed model and all the columns are characters and needs to have a lot of analysis on this, the best suite model is happening to be the linear regression model under supervised machine learning.

Apart from testing the data-sets from linear regression model it also being tested using the decision-tree model to see the various different accuracy score and choose the best suite model to yield the most accurate score.

These different accuracy score helps us make a clear decision on which model to use for the machine and have a better prediction level on making choices.

VII. SOME HELPFUL TIPS

1. Implementing various Models

While training any machine it need to find the way of accuracy, which means it needs to be trained with different algorithm models which helps is to understand in wide perspective.

Since different models will provide different accuracy scales, you will have a clear understanding of what model to choose.

2. Data Splits

On splitting the data one should make sure to understand the data and make a clear decision on what should be the target and independent data.

This will have a clear statement of the accuracy score to have a accurate prediction.

VIII. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions-ref[6]. Creating the model with additional parameters such as Different Location, skills, percentages, and ranking in their entrance exams etc, can make it more flexible to the admission departments to have a clear decision. Hence by generating the decision-making parameters, this system ca be used for any admission prediction process by taking into consideration all given and desired criteria.

APPENDIX

1. Raw data(admission_pred.csv)
2. Admission Selection model
   a) Input data (admission_pred.csv)
   b) Source Code (admission_pred.ipynb)
   c) Output (student_pred.rar)

REFERENCES

[1] athena.ecs.csus.edu [Online]
Available : http://athena.ecs.csus.edu/~pateljd/images/Admission_prediction_system.pdf

[2] towardsdatascience [Online]
Available : https://towardsdatascience.com/predicting-ms-admission-afbad9c5c599

[3] questia [Online]
Available : www.questia.com

[4] cs.utexas [Online]
Available : www.cs.utexas.edu

[5] EDMC [Offline]
Available : Student Paper

[6] ijsrp [Online]
Available : www.ijsrp.org