

University Admission Prediction using Machine Learning

Kruthika CS^{1*}, Apeksha B², Chinmaya GR³, Madhumathi JB⁴, Veena MR⁵

¹Department of CS&E.BIET, Davanagere. Karnataka, India

*Corresponding author: Kruthika C S, Department of CS&E.BIET, Davanagere. Karnataka, India ;Email: krithikac625@gmail.com

Received date: July 17, 2021; Accepted date: October 6, 2021; Published date: October 16, 2021

Citation: Kruthika CS, Apeksha B, Chinmaya GR, Madhumathi JB, Veena MR (2021) University Admission Prediction using Machine Learning. Am J Glob J Res Rev Vol: 8 No:7.

Abstract

In the present schooling world there are numerous quantities of understudies who need to seek after Higher training in the wake of Engineering or any Graduate certification course. Advanced education in the sense, a few groups need to do MTech through GATE or through any Educational Institute Entrance Examination and a few groups need to do MBA through CAT or through any individual Educational Institute Entrance Examination and a few groups need to do Masters in abroad colleges. Understudy confirmation issue is vital in Educational Institutions. We are addressing AI models to anticipate the opportunity of an understudy to be conceded to a Master's program. This will help understudies to know ahead of time in the event that they get an opportunity to get acknowledged. The Machine learning models are Linear relapse, Decision tree regressor and Random Forest regressor. Investigations show that the Linear Regression model outperforms different models.

Keywords: Linear regression; Admission prediction; Machine learning

Introduction

The world business sectors are growing quickly and constantly searching generally advantageous information and experience among individuals. Youthful specialists who need to hang out in their positions are continually searching for Higher degrees that can help them in working on their abilities and information. Thus, the quantity of understudies applying for Graduate examinations has expanded in the last decade. One of the principle concerns is getting conceded to their fantasy University. It's seen that understudies actually decide to get their schooling from universities that are known Universally. What's more, with regards to international alumni, the United States of America is the primary inclination of most of them. With most incredibly famous universities, Wide assortment of courses accessible in each order, exceptionally authorize instruction and educating programs, understudy grants are accessible for international understudies [1].

As per gauges, there are in excess of 10 million international understudies enlisted in more than 4200. Universities and

Colleges including both private and public across the United States. Generally, number of understudies concentrating in America are from Asian nations like India, Pakistan, Srilanka, Japan and China. They are picking America as well as UK, Germany, Italy, Australia and Canada. The quantity of individuals seeking after higher investigations in these nations are quickly expanding. The foundation justification the understudies going to abroad Colleges for Masters is the quantity of open positions present are low and number of individuals for those positions are exceptionally high in their separate nations. This moves numerous understudies in their calling to seek after Postgraduate investigations. It is seen that there is a significant huge number of understudies from Universities in the USA seeking after Masters in the field of Computer Science, the accentuation of this exploration will be on these understudies. Numerous schools in the U.S. follow comparative prerequisites for understudy affirmation. Schools consider various variables, for example, the positioning on fitness appraisal and scholastic record audit. The order over the English language is determined based on their exhibition in the English abilities test, for example, TOEFL and IELTS. The entrance advisory board of universities takes the choice to endorse or reject a particular up-and-comer based on the general profile of the candidate application. The dataset taken in this undertaking is identified with instructive area. Confirmation is a dataset with 400 lines that contains 7 distinct autonomous factors which are:

- Graduate Record Exam1 (GRE) score. The score will be out of 340 focuses.
- Trial of English as a Foreigner Language (TOEFL) score, which will be out of 120 focuses.
- University Rating (Uni.Rating) that demonstrates the Bachelor University positioning among different colleges. The score will be out of 5.
- Statement of direction (SOP) which is a record written to show the applicant's life, driven and the inspirations for the picked degree/college. The score will be out of 5 focuses.
- Letter of Recommendation Strength (LOR) which confirms the applicant proficient experience, fabricates validity, supports certainty and guarantees your ability. The score is out of 5 focuses.
- Undergraduate GPA (CGPA) out of 10.
- Research Experience that can uphold the application, like distributing research papers in gatherings, filling in as examination right hand with college teacher (either 0 or then again 1).

One ward variable can be anticipated which is possibility of affirmation, that is as per the input given will be going from 0 to 1.

Literature Survey

- One amazing work by Acharya et al. has looked at between 4 changed relapse calculations, which are: Linear Regression, Support Vector Regression, Decision Trees and Random Forest, to anticipate the opportunity of concede dependent on the best model that showed the least MSE which was multilinear relapse.
- Also, Chakrabarty et al. thought about between both linear regression and gradient boosting regression in foreseeing possibility of concede; call attention to that gradient boosting regression showed better outcomes.
- Gupta et al. fostered a model that reviews the alumni affirmation measure in American colleges utilizing AI procedures. The motivation behind this investigation was to direct understudies in tracking down the best instructive establishment to apply for. Five AI models were underlying this paper including Naïve Bayes, SVM (Linear Kernel), AdaBoost, and Logistic classifiers.
- Waters and Miikkulainen proposed an astounding article that aides in positioning graduation affirmation application as per the degree of acknowledgment and upgrades the presentation of inspecting applications utilizing measurable AI.
- S. Sujay applied linear regression to anticipate the shot at conceding graduate understudies in expert's projects as a rate. Be that as it may, no more models were performed.

Method Description

Data Collection

The way toward get-together information relies upon the sort of undertaking, for a ML project, real time information is utilized. The informational index can be gathered from different sources like a document, data set, sensor and different sources and some free informational collections from web can be utilized. Kaggle and UCI Machinelearning Repository are the storehouses that are utilized the most for information assortment for Machine learning models. Kaggle is quite possibly the most visited sites that is utilized for gathering informational collections [2,3].

Pre-processing

Information pre-processing is a cycle of cleaning the raw information i.e. the information is gathered in reality and is changed over to a perfect dataset. There are certain steps executed to change over the data into a little clean data collection and make it practical for examination, this piece of the interaction is called as information pre-processing.

The greater part of this present reality information is chaotic, as:

- Missing Data
- Noisy Data

• Inconsistent Data

A portion of the essential pre-handling strategies that can be utilized to change over crude data are:

- Conversion of Data
- Ignoring the missing qualities
- Filling the missing qualities
- Detection of exceptions
- Feature Extraction

At the point when the data to an algorithm is too enormous to ever be processed and it is suspected to be repetitive then it very well may be changed into a diminished arrangement of highlights. Deciding a subset of the initial features is called feature selection. The selected features are required to contain the significant data from the data, with the goal that the ideal assignment can be performed by utilizing this reduced representation rather than the total beginning data. Feature extraction includes reducing the number of assets needed to describe an enormous arrangement of information. When performing investigation of complex information one of the serious issues comes from the quantity of factors included. Examination with an enormous number of factors for the most part requires a lot of memory and calculation power, likewise it's anything but a grouping calculation to overfit to preparing tests and sum up inadequately to new examples. Highlight extraction is an overall term for techniques for developing blends of the factors to get around these issues while as yet portraying the information with adequate exactness. Many AI experts accept that appropriately improved element extraction is the way to successful model development [4,5].

Model Selection

Model determination is the way toward choosing one last AI model from among an assortment of applicant AI models for a training dataset. Model selection is a cycle that can be applied both across various kinds of models and across models of a similar sort arranged with various model hyper parameters.

Train and Test Data

For training a model we at first split the model into 2 segments which are 'training data' and 'Testing data'. The classifier is training utilizing 'training data set', and afterward tests the performance of classifier on inconspicuous 'test data set'.

Training set: The training set is the material through which the computer learns out how to deal with data. AI utilizes calculations to perform the training part. Training data set is utilized for learning and to fit the parameters of the classifier.

Test set: A set of unseen data utilized uniquely to evaluate the performance of a completely indicated classifier.

Evaluation Model

Assessment is an integral of the model development process. It assists with tracking down the best model that addresses the information and how well the picked model will function later

on. To work on the model hyper- boundaries of the model can be tuned and the exactness can be improved. Confusion matrix can be utilized to improve by expanding the quantity of genuine positives and genuine negatives. The output is predicted by investigating the test information as contribution alongside test information output and then the output is shown [6,7,8].

Interface

A web interface is built to take input and display an output. Flask language is used to build a web interface and pickle library is used to integrate both model and web page.

Data Visualization

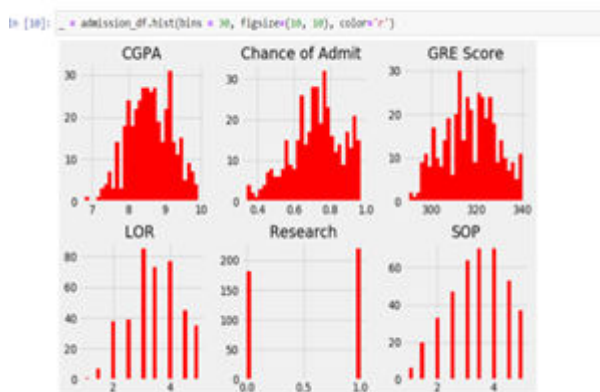


Figure1: Data visualization.

Data visualization is the representation of data in a graph, chart, or other visual format. Data visualization by using the histogram and analysing the data by the graph.

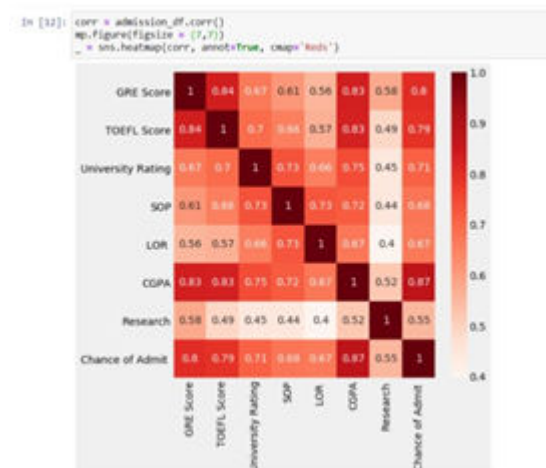


Figure2: Heatmap.

Data visualization by using the heatmap and analysing the data.

The top three features that affect the Chance of Admit are:

- CGPA
- GRE Score

- TOEFL Score
- CGPA vs Chance of Admit

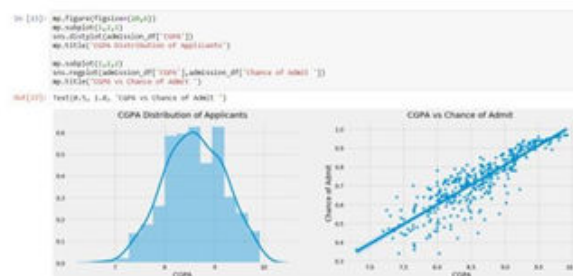


Figure3: CGPA vs Chance of Admit.

It appears as applicants CGPA has a strong correlation with their chance of admit.

GRE Score vs Chance of Admit

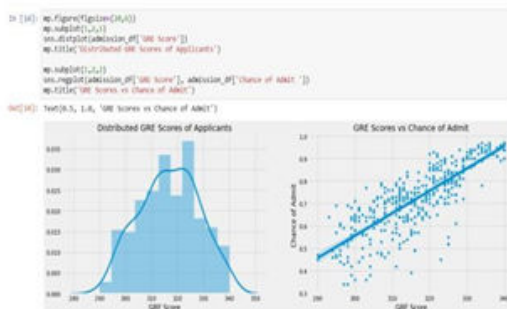


Figure4: GRE Score vs Chance of Admit.

GRE Score have a strong correlation with the chance of admission however not as strong as one's CGPA.

TOEFL Score vs Chance of Admit

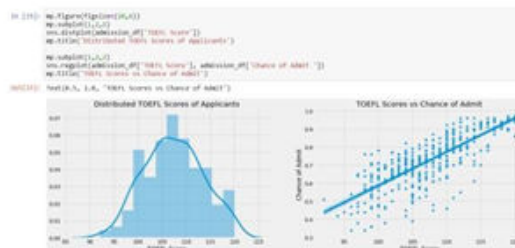


Figure5: TOEFL Score vs Chance of Admit

Research



Figure6: Research.

It appears to be most of candidates have research insight. In any case, this is the most un-significant element, so it doesn't make any difference to an extreme if a candidate has the experience or not [9,10,11].

Result

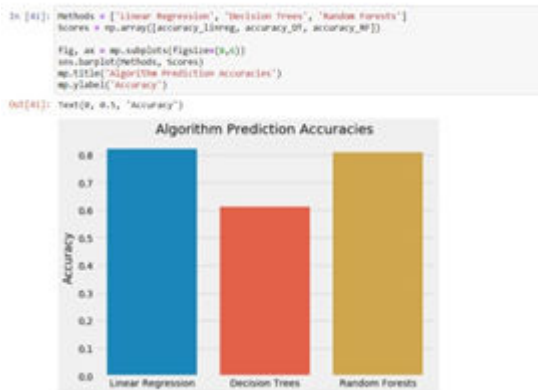


Figure7: Comparison graph.

In the above figure shows that the Linear Regression has more accuracy than different models. In the figure the Linear Regression shows 82% precision, Decision Tree shows 61% exactness and the Random Forest Regressor shows 80% exactness.

We can finish up with the Linear Regression model is productive and gives a superior outcome when contrasted with different models.



Figure 8: Result input image

The above figure shows the web interface of the Project. Here the users enter their exam results to check that he/she can get admit to that university or not, on the prediction percentage.



Figure9: Result predicted image

The above figure shows the web interface of the project. When the user enters the details followed by pressing the predict button, it shows the percentage of Chance of Admit. By giving the valid inputs we got 62% of admission chance to the university.

Conclusion

The primary objective of this work is to make a Machine Learning model which could be utilized by understudies who need to seek after their Education. Many AI algorithms were used for this examination. Linear Regression model contrasted with different models gives the best outcome. Understudies can utilize the model to survey their shots at getting induction into a specific University with a normal exactness of 82%. An ultimate objective of examination will be cultivated effectively, as the framework permits understudies to save the parcel of time and cash that they would spend on instructive guides and application charges for schools where they have less shots at getting affirmations. In future this module of expectation can be incorporated with module of robotized handling framework and different models like neural organization. Likewise, segregate investigation can be utilized independently or joined for upgrading dependability and precision forecast. At long last, understudies can have an open-source AI model which will assist the understudies with knowing their opportunity of entrance into a specific college with high exactness.

References

1. M. S. Acharya, A. Armaan, and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," Kaggle, 2018.
2. C. López-Martín, Y. Villuendas-Rey, M. Azzeh, A. Bou Nassif, and S. Banitaan, "Transformed k-nearest neighborhood output distance minimization for predicting the defect density of software projects," J. Syst. Softw., vol. 167, p. 110592, Sep. 2020.
3. M. S. Acharya, A. Armaan, and A. S. Antony, "A comparison of regression models for prediction of graduate admissions," ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc., pp. 1–5, 2019.

4. N. Chakrabarty, S. Chowdhury, and S. Rana, "A Statistical Approach to Graduate Admissions' Chance Prediction," no. March, pp. 145–154, 2020.
5. S. Sujay, "Supervised Machine Learning Modelling & Analysis for Graduate Admission Prediction," vol. 7, no. 4, pp. 5–7, 2020.
6. M. N. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Systematic ensemble model selection approach for educational data mining," *Knowledge-Based Syst.*, vol. 200, p. 105992, Jul. 2020.
7. Bruggink, T. H., and Gambhir, V. 1996. Statistical models for college admission and enrollment: A case study
8. for a selective liberal arts college. *Research in Higher Education* 37(2): 221–240.
9. Deerwester, S.; Dumais, S.; Furnas, G.; Landauer, T.; and Harshman, R. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6): 391–407.
10. Koh, K.; Kim, S.; and Boyd, S. 2007. An interior-point method for large-scale l1-regularized logistic regression. *Journal of Machine Learning Research* 8(8): 1519–1555.
11. Moore, J. S. 1998. An expert system approach to graduate school admission decisions and academic performance prediction. *Omega* 26(5): 659–670.