

Prediction of Admission Process for Gradational Studies using AI Algorithm

Saurabh Singhal, Ashish Sharma

Saurabh Singhal, Ashish Sharma

Department of Computer Engineering and Application, GLA University, Mathura

Department of Computer Engineering, GLA University, Mathura.

E-Mail: saurabh.singhal@gla.ac.in, ashish.sharma@gla.ac.in

Introduction

As there is advancement in Machine Learning, it is easy to develop an application accurately. There are various types of applications available for knowing whether we get seat in USA or not but mostly these are not reliable and not much effective and also building such applications is difficult but, Machine Learning provides us with several algorithms that are helpful to build a representation easily. The purpose of this work is to compare different Machine Learning algorithms and find out which algorithm is giving an accurate result for the available dataset. The algorithms we are going to use are Multi Linear Regression, Polynomial Regression and Random Forest. The input for these algorithms is GRE score, TOFEL score and CGPA of candidate. By using dataset we are going to train the representation and finally the output we are obtaining is the percentage of chance to get seat in reputed university.

Literature review

[1] Bayesian Networks were utilized by to make a choice emotionally supportive network for assessing the application put together by global scholars in the college. This representation was intended to anticipate the exhibition of the hopeful scholars by contrasting them and the presentation of scholar at present concentrating in the college and had a comparative outlined throughout their application. Right now on the present understudy's profile the representation anticipated whether the hopeful understudy ought to be allowed admission to the college. Since the correlations were made distinctly with the scholars who were at that point approved in the college and the information of the scholars who are deprived of confirmation were excluded from the examination this representation end up being less effective because of the issue of class lopsidedness.

[2] (Mishra and Sahoo (2016)) looked into from a college perspective to anticipating the probability of the scholar trying out the college after they have verified regarding courses in the college.

ABSTRACT

In the present time there are plenty of scholars seeking after their instruction away from their nations of origin. The fundamental nation focused through these worldwide scholars is The United States of America. The popular of the universal scholars in the United States of America are from India and China. With the expansion in the quantity of worldwide scholars concentrating in the USA, every candidate needs to confront extreme rivalry to get admission to their fantasy college. This work is to build up a framework utilizing AI algorithms, named it as Graduate Admission Prediction (GAP). GAP will assist the scholars by predicting the chance to get seat in Fantasy College. This paper compares and recognizes which AI algorithm is going to give precise outcome. A straightforward UI will be created for clients to get to the framework.

Keywords: Multi Linear Regression, Polynomial Regression, Random Forest.

Copyright

© 2020 The Author(s). This is an openaccess article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

They utilized the K-Means calculation for bunching the scholars depending on various components like criticism, family pay, family occupation, guardians capability, inspiration, and so forth to anticipate the scholars will enlist at the college or not. Contingent on the likeness of the traits between the scholars that are assembled into groups and choices are completed. The goal of the representation to build the enrolment of the scholars in the college.

[3] (Eberle et al. (n.d.)) utilized AI and prescient displaying to build up a representation to assess the confirmation approaches and benchmarks in the Tennessee Tech University. A notable variant of the C4.5 calculation, J48 was utilized to make the representation. Like the representations referenced above they utilized the various components of the understudy outlined to assess the odds of their admittance to the college. The representation functioned admirably in anticipating the genuine positive situations are the understudy had a decent outlined to make sure about the affirmation, yet it bombed in effectively recognizing the genuine negatives on account of which understudy that doesn't fulfill the characterized criteria.

[4] In look into directed by (Jamison (2017)) the yield of school affirmation was anticipated utilizing AI procedures. Yield rate can be characterized as the rate at that the scholars are been allowed affirmation by the college enlist for the course. Numerous AI calculations like Random Forest, Logistic Regression and SVM were utilized to make the representation.

Methodology

The below diagram is common for any type of algorithms those are the steps to follow while building a representation:

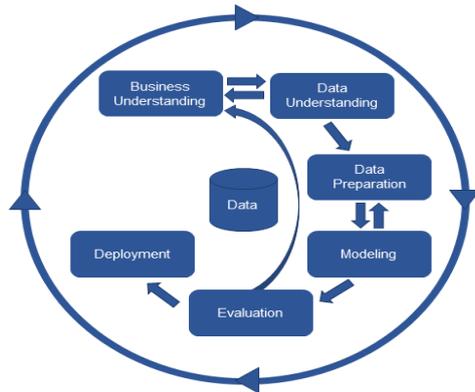


Fig 1: Architecture

Multi Linear Regression

It is An assessable method that uses a couple educational variables should anticipate the consequence of a response variable. That target about different straight regressions may be with demonstrating that straight association the middle of that self-sufficient variable and the dependent variable. Various straight relapse (MLR) is used choose a numerical association Around various unpredictable factors. The representational makes a association Likewise An straight line (linear) that best approximates every last one of individual majority of the information focuses.

At those point of convergence of the various straight relapse examinations may be those undertaking from claiming fitting an single transport through An scatter plot. Every last one of that's only the tip of the iceberg unequivocally the different straight relapse fits a accordance through An multi-dimensional space for data focuses.

The below diagram shows how the algorithm take best fit line Among all these lines which line have more data points nearly that is the best fit line.

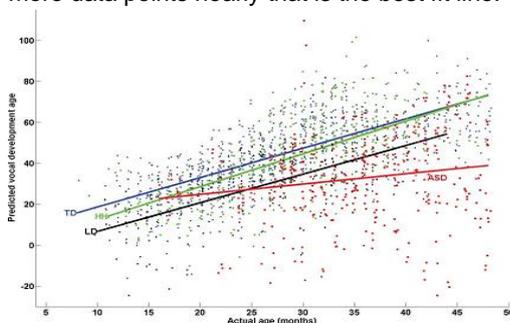


Fig 2: Multi Linear Regression

Polynomial Regression:

When there is chance for scattering of data a Multi Linear Regression (a straight line through all information focuses) may not work well, it may be perfect for Polynomial Regression. Polynomial Relapse, as straight relapse, utilizes the connection between the factors x and y to locate the most ideal approach to draw a line through the information focuses. Polynomial relapse will be An sort of Multi straight relapse the place the association between those self-sufficient variable x Also ward variable y is showed by An farthest breaking point polynomial. Polynomial relapse fits An nonlinear association the middle of the estimation about x and the comparing prohibitive imply of y, implied. $E(y |x)$.

The below diagram shows how our algorithm works Where ever the data points are available through that line passes non linearly and covers almost all the data.

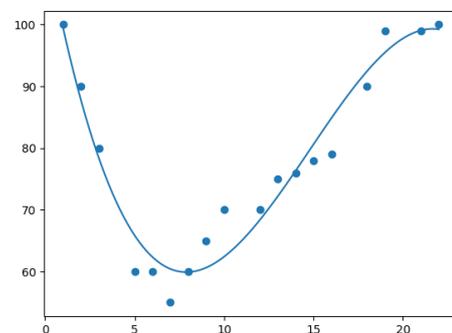


Fig 3: Polynomial Regression

Random Forest:

Random Forest is nothing but collection of decision trees. Every data what we are giving as an input is going to convert as one decision tree. Thus, several decision trees are constructed for single input data and we are going to collect all outputs from all decision trees and the highest counted output is considered. The Random Forest is a representation made up of many decision trees rather than just simply averaging the prediction of trees (which we could call a "forest").

The below diagram shows how this algorithm works At first our given input is going to form as decision trees and from each tree decisions are collected and finally which count is more is obtained as output.

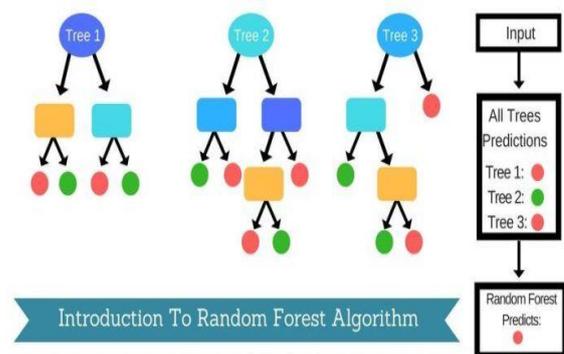


Fig 4: Random Forest

To implement above three representations we have to follow the following steps:

Step 1: Importing Libraries

Libraries required are pandas, NumPy and Multi-linear Regression classifier for prediction. Pandas are utilized for performing operations around information frames. Further, All the more utilizing NumPy, we will perform necessary scientific operations.

- **Numpy:**
Numpy is a Python package that stands for Numerical Python. It is the center library for logical processing, that holds an incredible n-dimensional display object, gives devices to coordinating C, C++, and so forth. It is additionally helpful in linear variable based linear algebra, random number capability, and so on.
- **Pandas:**
Pandas are a Python library generous greater information control and investigation devices utilizing its incredible information structures. Python with Pandas is utilized in a wide scope of fields.

Step 2: Reading the dataset

To pick the right variables, a basic understanding of the dataset is enough to know that data is relevant, high quality, and adequate volume. As part of our representation building efforts, we are working to select the best predictor variables for our representation.

Reading a CSV file:

```
df= pd.read_csv(body)
```

Step3: Training and Test Sets: Splitting Data

The dataset should divide into two subsets:

- Training set—a set to train a representation.
- Test set—a set to test the trained representation.

Splitting the data should meet the subsequent two circumstances:

- It is huge sufficient to give up statistically significant outcomes.
- It is illustrative of the informational index all in all. At the end of the day, don't pick a test set with unexpected attributes in comparison to the preparation set.

```
X = df.iloc[:, :-1].values
Y = df.iloc[:, 8].values
```

Sci-kit-learn give a scope of regulated and unaided learning calculations by means of a predictable interface in Python. It is endorsed under a tolerant disentangled BSD authorize and is disseminated below frequent Linux conveyances, authorize scholastic and business use. The library is depended upon the SciPy (Scientific Python) that must be initiated prior to we can utilize scikit-learn.

```
from sklearn.cross_validation import train_test_split
```

Step 4: Importing the algorithms from sklearn

```
from sklearn.linear_representation import Multi
- Linear Regression
```

Multi Linear Regression fits a linear representation with coefficients $w = (w_1, \dots, w_p)$ to limit the remaining total of squares between the watched focuses in the dataset, and the objectives anticipated by the linear guess.

Step 5: Create an object for our representation.

Step6: Fitting the trained data to our algorithm.

Step7: Calculating the test output.

Results and Discussions

For data representation the proposed work uses the Multi Linear Regression representation, polynomial Regression representation, Random Forest Regression representation and obtained an accuracies of 73%, 64% and 94% respectively.

The visualization technique we have used is Node-RED. Node-RED is a programming tool for wiring together equipment gadgets, APIs and online administrations in new and fascinating ways. It gives a program based manager.

Since mid 2014, the versatile has overwhelmed the PC (work area/PC) as the main gadget used to explore the Net. Alongside the versatile, numerous other compact gadgets that associate with the Internet have likewise begun multiplying at an extremely brisk rate. These days, the vast majority of us convey or have in any event one Internet-based gadget and portable. Along these lines, the Internet of Things presently doesn't just mean distinctive 'things' yet has developed into 'intelligent things' which have locally available calculation and system associations. Above all, they can detect nature around us and, appropriately, act intelligently. These gadgets are currently referred to as associated gadgets, shrewd items or the Web of Things.

After entering the marks in the respective column and by pressing submit we are going to obtain a chance to get a seat. The chance is in the form of points, If chance is near to 1 we have high chance, if chance is far to 1 we have low chance and if the chance is negative there is no chance to get seat.

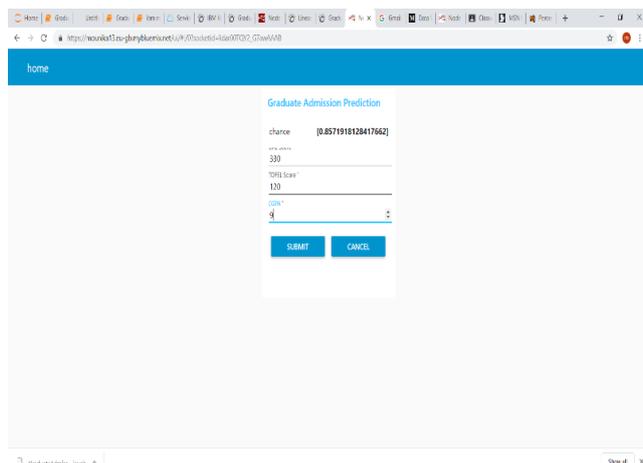


Fig 5: User Interface

The comparison of algorithms is shown in the graph below. X axis represents type of algorithm and y axis represents accuracy percentage of algorithm on the dataset.

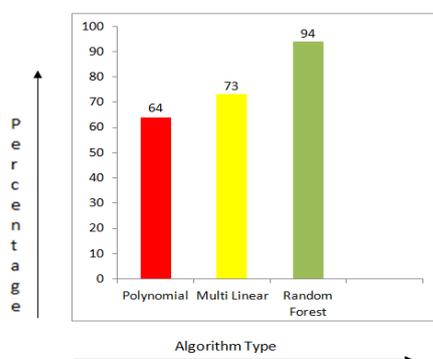


Fig 6: Comparison of Algorithms

Conclusion

Graduate Admissions Prediction representation is build by using different types of algorithms. Same dataset is trained for the three algorithms and by training the representation we obtained the results. Finally, Random Forest is the best algorithm to provide accurate result for the dataset that is available. The accuracy level we obtained using Random Forest is 94%, which is very encouraging. A user interface to interact with scholars to see the result is created using Node-RED.

Future Scope

From the proposed work we are able to identify only chance to get seat and we are not able to identify which university we are obtaining. So, in future we can develop a representation, which gives us a list of universities in which we can obtain admission.

References

1. Thi, N., Hien, N. and Haddawy, P. (2007). A Decision Support System for Evaluating International Scholar Applications, pp. 1 - 6.
2. Mishra, S. and Sahoo, S. (2016). A Quality-Based Automated Admission System for Educational Domain, pp. 221 - 223.
3. Eberle, W., Simpson, E., Talbert, D., Roberts, L. and Pope, A. (n.d.). Using Machine Learning and Predictive Representation to Assess Admission Policies and Standards.
4. Jamison, J. (2017). Applying Machine Learning to Predict Davidson College's Admissions Yield, pp. 765-766
5. M. M. P. K. P. S. J. Heena Sabnani, "Prediction of Scholar Enrolment Using Data Mining Techniques," April 2018. [Online]. Available: <https://www.irjet.net/archives/V5/i4/IRJET-V5I4408.pdf>.
6. J. Brownlee, "Linear Regression for Machine Learning," [Online]. Available: <https://machinelearningmastery.com/linear-regression-for-machine-learning/>. [Accessed 25 March 2016].
7. A. Géron, Hands on Machine Learning with Scikit Learn and Tensorflow, O'Reilly Media, 2017.
8. R. S. Brid, "Decision Trees," Greyatom, 26 Oct 2018. [Online]. Available: <https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>.
9. Kumar, Manoj, and Ashish Sharma. "Mining of data stream using "DDenStream" clustering algorithm." 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE). IEEE, 2013.
10. Sharma, Ashish, Anant Ram, and Archit Bansal. "Feature Extraction Mining for Student Performance Analysis." Proceedings of ICETIT 2019. Springer, Cham, 2020. 785-797.
11. Sharma, Ashish, and Dhara Upadhyay. "VDBSCAN Clustering with Map-Reduce Technique." Recent Findings in Intelligent Computing Techniques. Springer, Singapore, 2018. 305-314.
12. Sharma, Ashish, Ashish Sharma, and Anand Singh Jalal. "Distance-based facility location problem for fuzzy demand with simultaneous opening of two facilities." International Journal of Computing Science and Mathematics 9.6 (2018): 590-601.
13. Ram, Anant, et al. "A density based algorithm for discovering density varied clusters in large spatial databases." International Journal of Computer Applications 3.6 (2010): 1-4.
14. Kulshrestha, Jagrati, and Anant Ram. "An Analytical Study of the Chain Based Data Collection Approaches." 2019 4th International Conference on Information Systems and

- Computer Networks (ISCON). IEEE, 2019.
15. Agarwal, Rohit, A. S. Jalal, and K. V. Arya. "A review on presentation attack detection system for fake fingerprint." *Modern Physics Letters B* 34.05 (2020): 2030001.
 16. Mishra, Ayushi, et al. "A robust approach for palmprint biometric recognition." *International Journal of Biometrics* 11.4 (2019): 389-408.
 17. Singh, Anshy, Shashi Shekhar, and Anand Singh Jalal. "Semantic based image retrieval using multi-agent model by searching and filtering replicated web images." *2012 World Congress on Information and Communication Technologies*. IEEE, 2012.
 18. Shekhar, Shashi, et al. "A WEBIR crawling framework for retrieving highly relevant web documents: evaluation based on rank aggregation and result merging algorithms." *2011 International Conference on Computational Intelligence and Communication Networks*. IEEE, 2011.
 19. Varun K L Srivastava, N. Chandra Sekhar Reddy, Dr. Anubha Shrivastava, "An Effective Code Metrics for Evaluation of Protected Parameters in Database Applications", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 8, No.1.3, 2019. doi.org/10.30534/ijatcse/2019/1681.32019
 20. N Chandra Sekhar Reddy, Dr. Purna Chandra Rao Vemuri, Dr. A Govardhan, Ch. Vijay, "An Empirical Study On Feature Extraction Techniques For Intrusion Detection System", *Journal of Advanced Research in Dynamical and Control Systems*, Vol. 9. Sp– 12 / 2017.