



Machine Learning-based Prediction and Possibility for University Admission based on Student's Profile.

By

MD. Al-Mamun

181-35-300

A thesis submitted in partial fulfillment of the requirement for the degree of
Bachelor of Science in Software Engineering

Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY

Fall – 2021

APPROVAL

This thesis titled on “Admission prediction using machine learning”, submitted by **Md. Al Mamun, 181-35-300** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS



Chairmn

Dr. Imran Mahmud
Associate Professor and Head
Department of Software
Engineering Daffodil International
University



Internal Examiner 1

Kaushik Sarker
Assistant Professor
Department of Software
Engineering Daffodil International
University



Internal Examiner 2

Md. Shohel
Arman Senior
Lecturer
Department of Software
Engineering Daffodil International
University



External Examiner

Md. Fazle
Munim
Technology
Expert
Access to Information (a2i) Programme

DECLARATION

It hereby declares that this thesis has been done by us under the supervision of **Kaushik Sarker**, Assistant Professor & Associate Head, Department of Software Engineering, Daffodil International University. It is also declared that neither this thesis nor any part of this has been submitted elsewhere for the award of any degree.

Mamun

Md. Al-Mamun

Student ID: 181-35-300

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

Certified by:

Kaushik Sarker

Kaushik Sarker

Assistant Professor & Associate Head,

Department of Software Engineering

Faculty of Science & Information Technology

ACKNOWLEDGEMENT

First of all, we are grateful to the Almighty Allah for giving us the ability to complete the final thesis.

We would like to express our gratitude to our supervisor **Kaushik Sarker** for the consistent help of my thesis and research work, through his understanding, inspiration, energy, and knowledge sharing. Her direction helped us to find the solutions to research work and reach our final theory.

We would like to express my extreme sincere gratitude and appreciation to all of our teachers of the **Software Engineering** department for their kind help, generous advice, and support during the study.

We are also express our gratitude to all of our friends, senior, junior who, directly or indirectly, have lent their helping hand in this venture.

Last but not the least, we would like to thank our family for giving birth to us at the first place and supporting me spiritually throughout my life.

Table of Contents

Sl	Description	Page
1	APPROVAL	i
2	DECLARTION	ii
3	ACKNOWLEDGEMENT	iii
4	ABSTRAC	vii
5	CHAPTER 1 INTRODUCTION	1
6	CHAPTER 2 LITERATURE REVIEW	2-3
7	CHAPTER 3 METHODOLOG 3.1 Linear Regression 3.2 Random forest 3.3 Decision tree 3.4 Artificial Neural Network	4-7
8	CHAPTER 4 RESEARCH MODEL	8
9	CHAPTER 5 ABOUT DATA SET 5.1 Data description 5.2 Plotting Histogram 5.3 Features correlation	9-12
10	CHAPTER 6 RESULTS DISCUSSION	13-14
11	CHAPTER 7 CONCLUSION AND FUTURE WORK	15
12	REFERENCES	16-17

List of Tables

Sl	Description	Page
1	Tab.1. describe data set about all input	9
2	Tab.2. Plotting Histogram	10-11
3	Tab.3. shows the result from a different model.	13

List of Figures

Sl	Description	Page
1	Fig.1. how work random forest	5
2	Fig. 2. Artificial neural network model loss after train	7
3	fig. 3 . Research Model process step by step	8
4	Fig. 4. Correlation between features	12
5	Fig. 5. Visualize the percentage of the whole model outcome.	14

ABSTRACT

The decision of which university to go to for postgraduate studies is quite challenging for students. Many mentors give guides for a fee. It does not help the majority of students. There are a lot of people who have been duped. Despite the fact that I have a similar difficulty while considering moving overseas for my higher education. Based on my profile, I look into a variety of institutions. However, based on my profile, I am unable to locate the ideal university. With the number of universities and subjects growing by the day, deciding which institution is ideal for a well-profiled student is becoming increasingly difficult. To address this issue, Given the student's profile I propose a machine learning-based method that compares various regression algorithms such as Artificial neural network, decision tree, Random Forest, linear regression. This paper compares and contrasts different models to determine which one is the most accurate.

Keywords- Admission Prediction, Linear Regression, Artificial Neural Network, Radom Forest, Decision tree

CHAPTER 1

INTRODUCTION

Graduation students are rising all the time, and it is becoming increasingly tough to get into a master's program. It's becoming increasingly difficult. Going overseas for my master's degree is a dream of mine. So, on occasion, I seek a master's degree program but am unable to locate one that is perfect for my needs. As a result, I consider the necessity to take action.

Basically, when it comes to applying to master's programs, prospective graduate students typically confront several hurdles. It is currently difficult for a student to select his or her desired university for graduation. Given everything, deciding which university is best for him is challenging. even though I face this kind of problem when I think about going abroad for my higher study. I research lots of universities based on my profile. but I don't find out the quintessential university based on my profile. day by the increasing university and subject and so it's very tough to decide which university is best for a well profile based student. Using data from Kaggle, this paper presents the implementation of four machine learning algorithms as a method of predicting potential university ratings to admit the students.

we compare various regression algorithms such as Artificial neural network decision tree, Random Forest, linear regression. This paper compares and contrasts different models in order to determine which one is the most accurate.

CHAPTER 2

LITERATURE REVIEW

Graduating students are becoming increasingly competitive for admission to the next level of master's programs since a growing number of students are unaware of the requirements. They squandered too much money on consulting counsel [1]. This paper used standards based on features like Research Experience, GRE, CGPA, Undergraduate University. not use other significant elements like standard operating procedure(sop), Letter of recommendation(LOR), Research [2]. In this study, we'll merely look at MSE to see how accurate it is. MSE or RMSE are useful for determining accuracy. However, if the accuracy is shown as a %, it will be more accurate. It is more suitable for real-time visualization [3]. Too many advisors and predictors are needed to recommend the best university for a student's profile. However, they are not always correct in their predictions. They make mistakes from time to time [4]. For model creation, using supervised machine learning algorithms. and acquire accuracy of 75% by using just graduation CGPA and physics GRE score. Higher variables, such as university rating, SOP, and LOR, are needed for more accuracy.[5]. It's a statistical method for predicting admission that calculates the R squared error, MSE, and RMES for a better understanding of the predictive value [6]. Classifying students into several groups for various university admissions methods. and analyzing a wide range of student data in order to forecast their university entrance [7]. Because this is not a classification model, they utilized semi-supervised learning and data mining, however for a better result, need to employ a deep learning method. In my paper, I employed ANN as a deep learning algorithm.[9]. The globe is advancing at such a quick pace that everything is moving on an autonomous basis in order to reduce time in the west and to move faster. in improving world are need qualified people for betterment . This application aids in the identification of the ideal student for the job based on her qualifications[10]. For a better outcome, utilize the supervised machine learning algorithm. For a better outcome, I combined supervised and deep learning algorithms[11].data mining technique[12].using

This paper used standards based on features like Research Experience TOEFL, GRE, CGPA, Undergraduate University. I think more characteristics such as Research Experience, LOR, and SOP should be added to improve prediction [13]. in this paper using deep neural network DNN.[14]. This work employs a machine learning method as well as EDM.[17]. Use machine learning-based framework for predicting student performance[20].

CHAPTER 3

Methodology

Background Study:

3.1 Linear Regression:

Linear Regression Algorithm is the most simplified and widely used algorithm to forecast the result. This algorithm is so helpful for statics and machine learning. It's work very well these two-part. Researchers are like the linear model. Because Linear delineations were pretty much simple. So it is a seductive model for every researcher. Basically, it's established on supervised learning. What adjectly linear regression?. The main objective is to obtain a relationship between an independent variable(X) and the dependent variable(y). Lots of the easiest linear models have just one independent and one dependent variable. In my case have more than one variable so I use multiple linear algorithms.

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^m (\sigma(\theta^T \cdot \mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

$$y = c + b \cdot x,$$

The equation for a linear regression line is $Y = a + bX$. here b is the slope, x means the feature variable and a is a Y-intercept y is mention the target variable value. make sure to find out the value of θ that decrease the RMSE by the Equation MSE cost function for a Linear.

$$\text{MSE}(\mathbf{X}, \mathbf{h}_\theta) = \frac{1}{m} \sum_{i=1}^m (\theta^T \cdot \mathbf{x}^{(i)} - y^{(i)})^2$$

here x means the instance's feature vector, h_θ = hypothesis function using the model parameters θ , m = number of sample in dataset θ^T is the transpose of θ , $\theta^T \cdot \mathbf{x}^{(i)}$ = is the dot product of θ^T and $\mathbf{x}^{(i)}$, and y = expected value.

3.2 Random forest:

In a random forest, We need to investigate the ensemble approach. Ensemble means combining multiple models.

For prediction, this multiple model (Ensemble) is better than an individual model. This model combines the result of several decision trees on different subsets of a dataset and averages them to increase the dataset's predicted accuracy.

Figure 1: Random Forest

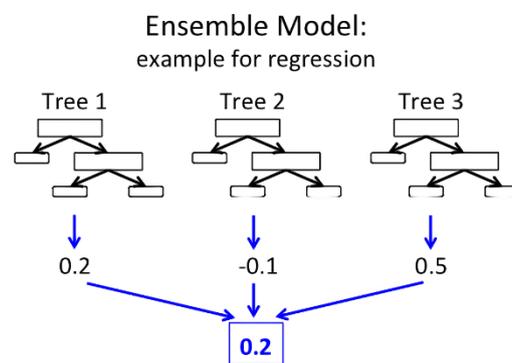


Fig. 1. how work random forest

3.3 Decision tree:

The decision tree resembles a tree structure and provides a final choice. That is why call decision trees. this algorithm makes for classification or regression models. It gradually Breaks down a dataset into smaller and smaller sections while also developing an associated decision tree. A tree containing decision nodes and leaf nodes is the end result. now I show the formula that minimizes the cost function.

$$J(K, L_k) = \frac{m_{left}}{m} G_{left} + \frac{m_{right}}{m} G_{right}$$

Where

$\left\{ \begin{array}{l} G_{left/right} \text{ measures the impurity of the left/right subset.} \\ m_{left/right} \text{ is the number of instances in the left/right subset.} \end{array} \right.$

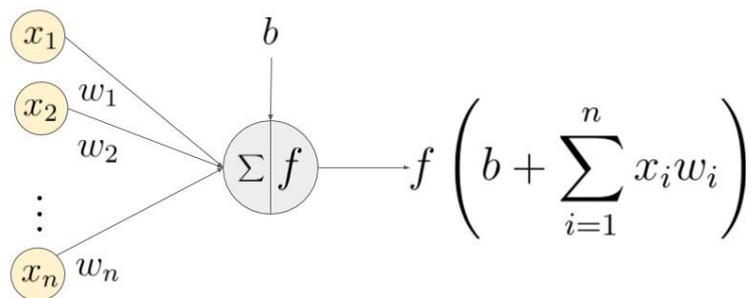
To train a linear regression model make sure to find out the value of θ that decrease the RMSE by the Equation MSE cost function for a Linear.

$$\text{MSE}(\mathbf{X}, \mathbf{h}_\theta) = \frac{1}{m} \sum_{i=1}^m (\theta^T \cdot \mathbf{x}^{(i)} - y^{(i)})^2$$

here \mathbf{x} means the instance's feature vector, \mathbf{h}_θ =hypothesis function using the model parameters θ , m = number of sample in dataset θ^T is the transpose of θ , $\theta^T \cdot \mathbf{x}^{(i)}$ = is the dot product of θ^T and $\mathbf{x}^{(i)}$, and y = expected value.

3.4 Artificial Neural Network:

The artificial neural network is a computational algorithm. Artificial neural network based on the architecture of the biological neural network. So Artificial is called a neural network. IT works like the human brain include neurons that are coupled to one another in various levels of the networks. Nodes are the name for these neurons.



Here showing the input(x_1 - x_n),their corresponding weights(w_1 - w_n),a bias(b) and the activation function f applied to the weighted of the inputs.

Figure 1: Artificial neural network

Model loss after train.

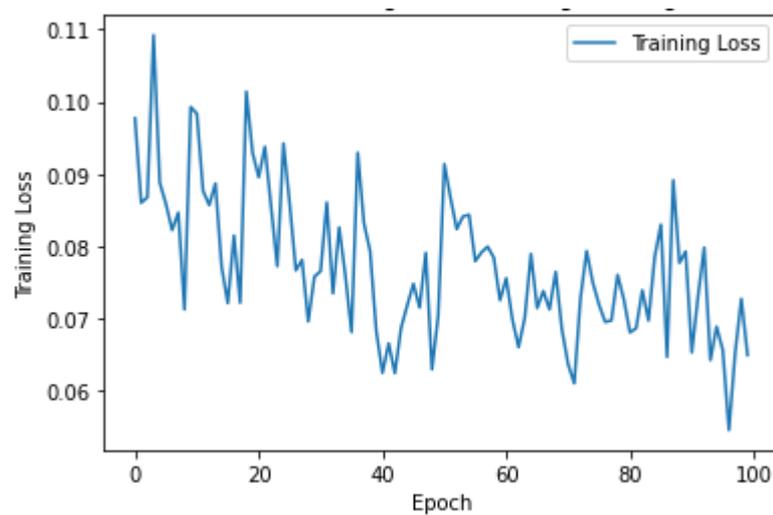


Fig. 2. Artificial neural network model loss after train

CHAPTER 4

Research Model

This work used four Regression models Artificial Neural Network, Decision Tree, Random Forest, and Linear Regressor. I utilized this model to better forecast the likelihood of being admitted. The accuracy of the models was used to assess their performance. I use google colab to run this work. Then I preprocess the data set and visualization the data set. And train the model and used to predict the chance of admission. the select the best-curated model with less error. . I calculated RMSE to choose the best model. Then compare my model Artificial Neural Network, Decision Tree, Random Forest, and Linear Regressor. In my model linear Regressor is the best model. Linear Regressor is better than another model. For linear Regressor, RMSE is 0.061. That is minimum than another model.

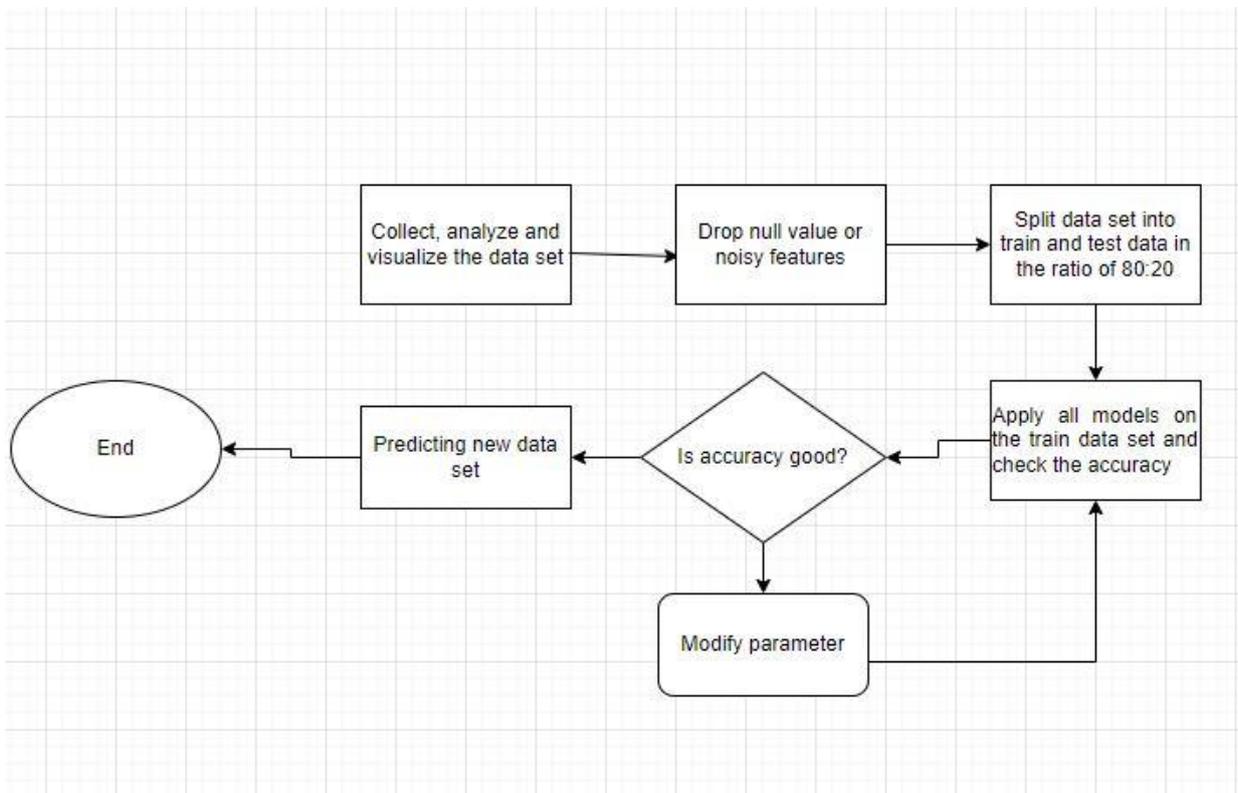


fig. 3 . Research Model process step by step

CHAPTER 5

About data set

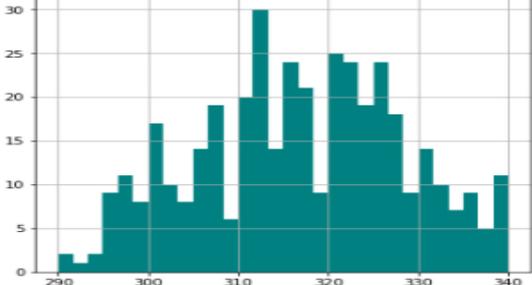
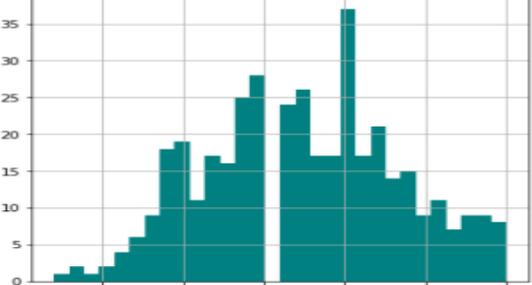
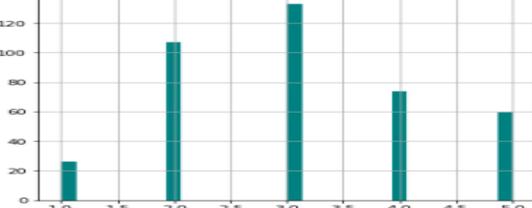
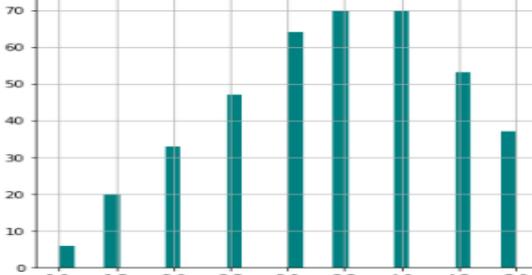
5.1 Data description

Table 1: describe data set about all input

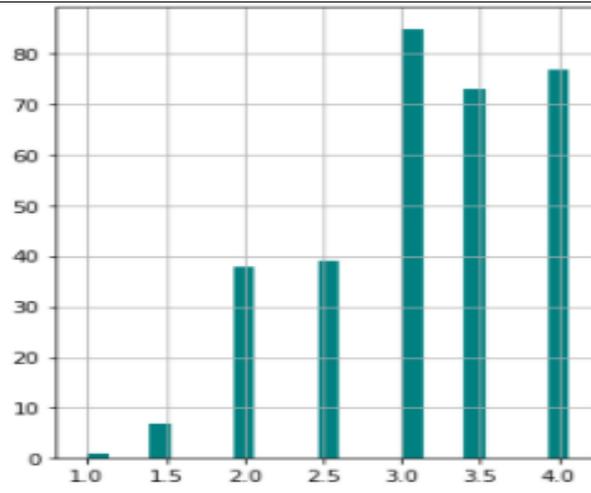
Column	description
GRE Score (out of 340)	The graduate record examination (GRE) is a standardized exam for measuring student's aptitude for abstract thinking
TOEFL Score (out of 120)	Test of English as a Foreign Language (TOEFL) score
University Rating (out of 5)	is the rankings of the higher education institute
SOP (out of 5)	A Statement of Purpose is an essay of purpose of applying to a specific course in a particular university.
LOR (out of 5)	A Letter of Recommendation (LOR) from a professional who has taught a student.
CGPA (out of 10)	CGPA is the average grades obtained by a student in all the semesters
Research (either 0 or 1)	"Research Experience" means any academic research activity.
Chance of Admit (ranging from 0 to 1)	Is the likelihood of a student's chance to be admitted to a university. 1 means great possibility, 0 means no chance of admission.

5.2 Plotting Histogram

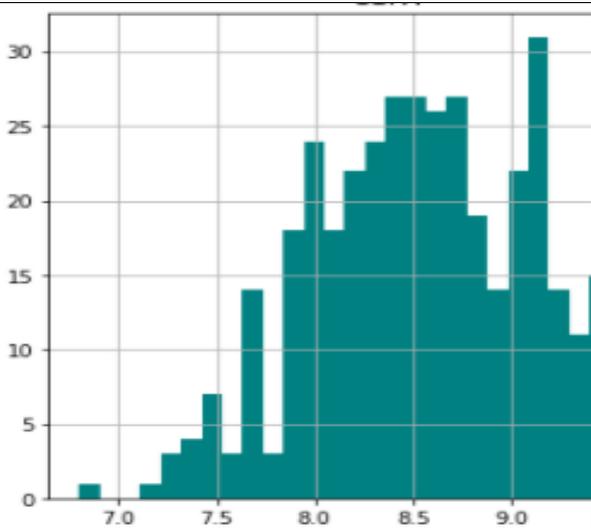
Table 2. for better understanding the data set.

Attribute	Histogram
<p style="text-align: center;">GRE Score (out of 340)</p>	
<p style="text-align: center;">TOEFL Score (out of 120)</p>	
<p style="text-align: center;">University rating (out of 5)</p>	
<p style="text-align: center;">SOP (out of 5)</p>	

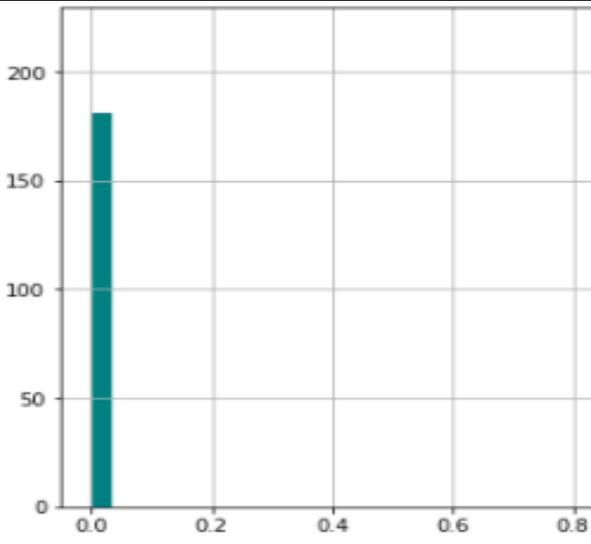
LOR (out of 5)



CGPA (out of 10)



Research (either 0 or 1)



5.3 Features correlation

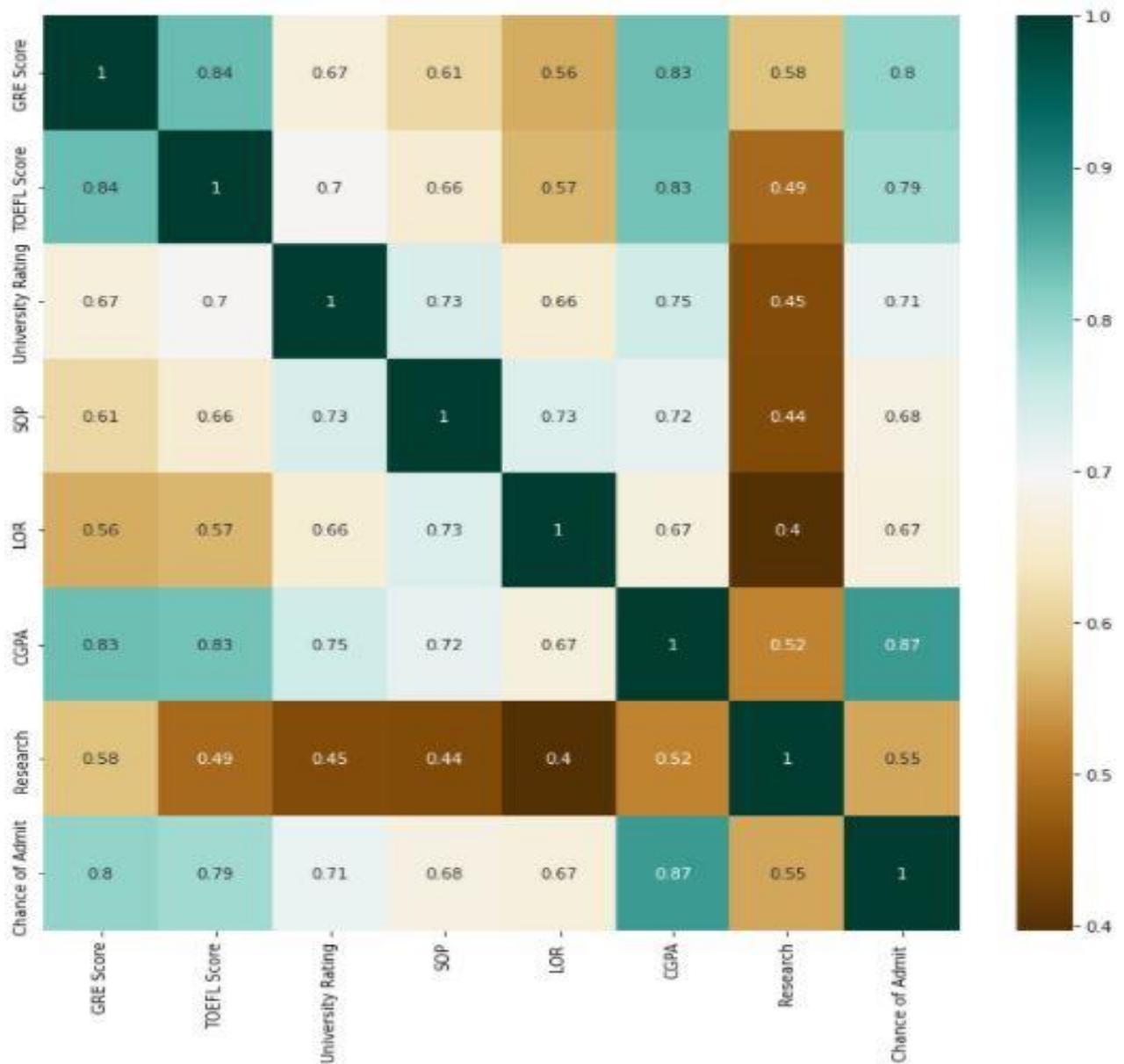


Fig. 4. Correlation between features

Student have hijer GRE Score and CGPA and also had TOEFL. Correlation matrices show that these characteristics are significantly connected.

CHAPTER 6

RESULTS DISCUSSION

I Used Four models to predict better accuracy, using an Artificial Neural Network, Decision Tree, Random Forest, and Linear Regressor. we build this model for better accuracy. the different model gives different accuracy. then we compare which model gives the best accuracy. here artificial neural network gives 77.47% and decision tree gives 50.65% and random forest gave 79.10%. For better calculation, I calculate RSME that's helpful for abject accuracy. here I just show minimal RSME value it's from the Linear Regression model. for all overviews we decided Linear Regression is the best model. And linear Regression gives the highest accuracy 80.30%. And linear Regressor smallest RMSE (0.061). and other models show Lowe accuracy as shown in the table.

Table 3: shows the result from a different model.

Model name	accurecy
Artificial Neural Network	<pre>print("ANN model accuracy: {:.2f} %".format(ANN_accuracy*100)) 2/2 [=====] - 0s 6ms/step - loss: 0.2253 ANN model accuracy: 77.47 %</pre>
Decision Tree	<pre>## Accuracy DecisionTree_accuracy = DecisionTree_model.score(X_test, y_test) print("Decision Tree model accuracy: {:.2f} %".format(DecisionTree_accuracy*100)) Decision Tree model accuracy: 50.65 %</pre>
Random Forest	<pre>## Accuracy RandomForest_accuracy = RandomForest_model.score(X_test, y_test) print("Random Forest model accuracy: {:.2f} %".format(RandomForest_accuracy*100)) Random Forest model accuracy: 79.10 %</pre>
Linear Regressor	<pre>## Accuracy LinearRegression_accuracy = LinearRegression_model.score(X_test, y_test) print("Linear Regression model accuracy: {:.2f} %".format(LinearRegression_accuracy*100)) Linear Regression model accuracy: 80.30 %</pre>

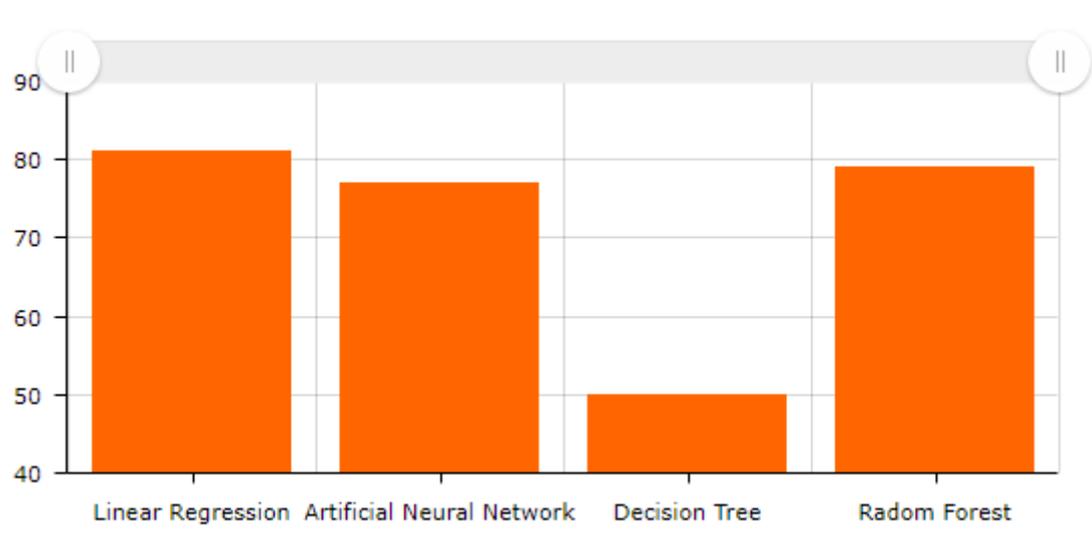


Fig. 5. Visualize the percentage of the whole model outcome.

CHAPTER 7

CONCLUSION AND FUTURE WORK

After finishing our work we can decide the best model is linear regressor based on accuracy. So we compare with another model Random forest, decision tree, and artificial neural network for selecting the model. I use this algorithm to see how to put all of your master's qualifications to good use forecast the possibility of admission in various values. linear regressor is gained the most accuracy 80.30% and a minimal RMSE value of 0.061 on the other hand, Other models give low accuracy compared with linear regression. . here artificial neural network gives 77.47% and decision tree gives 50.65% and random forest gave 79.10%.

so the main aim of this paper for create an application where students find out the best university based on her/his profile for a postgraduate degree.

Students' chances depend on him/her profile. the most prominent of which are GPA, GRE, and TOEFL, university rating, SOP, LOR, Research. I used a total of 7 features in this paper. but the best thing is more features give more accuracy. Other characteristics might include industry experience, internship experience. Also, as an extension to this work, a university suggestion for research interest can be produced with more research.

REFERENCES

- [1]. AlGhamdi, A., Barsheed, A., AlMshjary, H. and AlGhamdi, H., 2020, March. A machine learning approach for graduate admission prediction. In *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing* (pp. 155-158).
- [2]. Manley, J. and Krishnakumar, S., University Recommender System for Graduate Studies in USA.
- [3]. Sivasangari, A., Shivani, V., Bindhu, Y., Deepa, D. and Vignesh, R., 2021, April. Prediction Probability of Getting an Admission into a University using Machine Learning. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1706-1709). IEEE.
- [4]. Acharya, M.S., Armaan, A. and Antony, A.S., 2019, February. A comparison of regression models for prediction of graduate admissions. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)* (pp. 1-5). IEEE.
- [5]. Young, N.T. and Caballero, M.D., 2019. Using machine learning to understand physics graduate school admissions. *arXiv preprint arXiv:1907.01570*.
- [6]. Chakrabarty, N., Chowdhury, S. and Rana, S., 2020. A statistical approach to graduate admissions' chance prediction. In *Innovations in Computer Science and Engineering* (pp. 333-340). Springer.
- [7]. Fong, S. and Biuk-Aghai, R.P., 2009, December. An automated university admission recommender system for secondary school students. In *The 6th International Conference on Information Technology and Applications* (p. 42).
- [8]. Bag, A., 202100. A comparative study of regression algorithms for predicting graduate admission to a university.
- [9]. Zhao, Y., Lackaye, B., Dy, J.G. and Brodley, C.E., 2020. A Quantitative Machine Learning Approach to Master Students Admission for Professional Institutions. *International Educational Data Mining Society*.
- [10]. Waters, A. and Miikkulainen, R., 2014. Grade: Machine learning support for graduate admissions. *Ai Magazine*, 35(1), pp.64-64.
- [11]. Sujay, S., 2020. Supervised machine learning modelling & analysis for graduate admission prediction. *Int. J. Trend Res. Dev*, 7(4), pp.5-7.
- [12]. Sabnani, H., More, M., Kudale, P. and Janrao, S., 2018. Prediction of Student Enrolment Using Data Mining Techniques. *International Research Journal of Engineering and Technology (IRJET)*, 5(4), pp.1830-1833.
- [13]. Shrestha, K., 2017. *UNIVERSITY RECOMMENDATION AND ADMISSION PREDICTION SYSTEM (UniGrade)* (Doctoral dissertation, TRIBHUVAN UNIVERSITY).
- [14]. Goni, M.O.F., Matin, A., Hasan, T., Siddique, M.A.I., Jyoti, O. and Hasnain, F.M.S., 2020, December. Graduate Admission Chance Prediction Using Deep Neural Network. In *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (pp. 259-262). IEEE.
- [15]. El Guabassi, I., Bousalem, Z., Marah, R. and Qazdar, A., 2021. A Recommender System for Predicting Students' Admission to a Graduate Program using Machine Learning Algorithms.

- [16]. Ofori, F., Maina, E. and Gitonga, R., 2020. Using Machine Learning Algorithms to Predict Students'™ Performance and Improve Learning Outcome: A Literature Based Review. *Journal of Information and Technology*, 4(1), pp.33-55.
- [17]. Iqbal, Z., Qadir, J., Mian, A.N. and Kamiran, F., 2017. Machine learning based student grade prediction: A case study. *arXiv preprint arXiv:1708.08744*.
- [18]. Waters, A. and Miikkulainen, R., 2014. Grade: Machine learning support for graduate admissions. *Ai Magazine*, 35(1), pp.64-64.
- [19]. Qazdar, A., Er-Raha, B., Cherkaoui, C. and Mammass, D., 2019. A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, 24(6), pp.3577-3589.
- [20]. El Guabassi, I., Al Achhab, M., Jellouli, I. and El Mohajir, B.E., 2016, October. Recommender system for ubiquitous learning based on decision tree. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)* (pp. 535-540). IEEE.