ISSN 2278-3091
Volume 10, No.6, November - December 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse051062021.pdf
https://doi.org/10.30534/ijatcse/2021/051062021

Analysis and Prediction of Graduate Admissions Based on Pre-COVID and post-COVID Scenario

Sagnik Sarkar[1], Sriya I. Reddi[2], Manjula R[3]
[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, sagnik.sarkar2018@vitstudent.ac.in
[2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, sriyai.reddi2018@vitstudent.ac.in
[3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, rmanjula@vit.ac.in

Received Date : October 07, 2021 Accepted Date : November 09, 2021 Published Date : December 06, 2021

ABSTRACT

Graduate admissions is one of the events that attracts a lot of attraction from prospective students and universities alike. Be it the university conducting graduate admissions or an aspiring student; both yearn for a prediction system to aid in the process of selecting admits. On one hand, the university can get an insight on the probability of a student's admit thus aiding the graduate admissions office in their workload, and on the other hand the student can get a forecast on the chance of admit and can take preemptive decisions to facilitate the process. However, due to the COVID-19 pandemic, the graduate admissions has seen a slight change in paradigm. This change creates confusion among the related masses. A probing analysis on this change serves as a reference to act upon. In this study, prediction models are built with an extra parameter signifying whether a record in the dataset belongs to the COVID-19 pandemic period. Various models such as Logistic Regression, Decision Tree, Random Forest, Gaussian Naive Bayes and Artificial Neural Networks are used to determine the change in probability of admission due to the effect of the pandemic. All the models provide an accuracy score in the range of about 55% to 80%, with the Neural Network outperforming all the other models with a test accuracy score of 79.03%. The effect of the pandemic has caused an ambiguous response to various factors, but it can be stated the chances of admits of students have generally increased likely due to the lower number of applicants.

Key words : Neural Networks, Analysis, Graduate Admissions, COVID19

1. INTRODUCTION

Students often have multiple questions about universities whether they can get admission, scholarship, and accommodation. One of the main concerns is getting admitted to their dream university. It is observed that students still choose to obtain their education from universities that are well-known internationally. And when it comes to international graduates, the United States of America is the first preference for the majority of them. With most world-renowned colleges, wide variety of courses available in each discipline, highly accredited education and teaching programs, student scholarships, are available for international students. According to estimates, there are more than 10 million international students enrolled in over 4200 universities and colleges including both private and public across the United States. The number of people pursuing higher studies in these countries are rapidly increasing. The background reason for the students choosing to study in universities abroad, for Masters, is the number of job opportunities present are low and number of people for those jobs are very high in their respective countries. This inspires many students to pursue postgraduate studies for their respective fields.

This process is tedious and stressful and a lot of students pay counselors a lot of money to learn about the colleges they could get a guarantee admission into with their current academic and extra-curricular statistics. To help students cutback on the high fees they pay these counselors we propose to provide analysis based on various prediction algorithms. These algorithm will predict the chances of a student getting an admission in various colleges based on their test scores, work experience, and other considerable factors. Although datasets are available for this particular problem they haven't updated the information for 2020 and 2021. But due to COVID-19, numerous students deferred their admission to the next year. As a consequence the acceptance rates of colleges went down drastically to accommodate those students in the next batch, considering the limited number of seats allotted per year. To account for this change and to see how this can affect the admissions of the students in the upcoming years we want to scrape the data, for the admits and rejects from these two years (2020 and 2021), and analyze the admission decisions based on pre-COVID-19 and post-COVID-19.

2. LITERATURE SURVEY

Graduate Admissions have been looked at with interest in the recent years. The literature consists of various approaches dealing with this topic; from a psychological and analytical perspective to a predictive approach. The major focus of the recent literature include analytical and predictive methods harnessing Machine Learning [1] and other statistical methods [2] to predict the probability that an aspiring candidate getting accepted into a certain institution. This section deals with the discussion on various Machine Learning and Statistical methods used in recent works.

[1] introduces a Machine Learning based system GRADE (graduate admissions evaluator) which helps aid the admissions committee make the process of reviewing admission applications more efficient. This system is trained on the past admission decisions available in their department's database. This database contains student information and a binary label indicating whether the student had been admitted or not. The student information consists of a multi-dimensional vector representing various attributes of institutions previously attended, GPAs, Test scores, LORs, Area of research interest and preferred faculty advisor. This system uses an L1 - regularized logistic regression model. Based on the results obtained, the decision process is similar to the committee decisions.

[3],[4],[5], and [6] use numerous regression and statistical techniques for predictions and analyze their performance on graduate admissions dataset. The models compared include, Linear Regression, Support Vector Regression, Logistic Regression, KNN Classification, Naive Bayes Classification, Decision Trees and Random Forest. The attributes considered for each aspirant in the dataset include Letter of Recommendation, Statement of purpose and scores like GRE, TOEFL and undergraduate GPA. The evaluation metrics used for comparison and benchmarking in [3] are MSE (Mean Squared Error), RMSE (Root Mean Squared Error), mean Squared Log Error and R2 Score. It is concluded from the observations of the results obtained that, due to the linear nature of the data in the n-dimensional space, linear regression performs better than other regression methods. However, the results obtained by [6] and [5] indicate that the Random Forest Regressor and logistic regression models respectively perform better than the other algorithms. In comparison to Decision tree, Random Forest, Adaboost and SVC, [4] reports that Naive Bayes algorithm performs the best on the graduate admissions data to be used in a prediction system. [7] takes a similar approach and utilizes a gradient boosting regressor model to obtain a binary classification of probability of the admission of the student. R2 Score and other performance metrics similar to [3] have been used for comparison.

A Deep Learning approach has been taken by [8] for the prediction algorithm. The attributes of the data is similar to the previous papers. The dataset after normalization is passed to the Deep Neural Network Model consisting of 3 layers of neurons. Based on the training and analysis, an importance score of each attribute is obtained. The undergraduate GPA and GRE score are the most important features in the graduate admissions process. An R2 score of 0.8538 is obtained from the deep learning model. [9] also uses a deep learning model for prediction and compares the performance to the basic statistical methods. This represents the superior feature extraction and representation capability of neural networks in comparison with basic statistical models. [10] takes a stacked ensemble learning approach to increase the accuracy and overcome the variation in the data. A stacked ensemble of Deep Neural Networks is used for the prediction of the admissions. Ensemble Neural Networks outperform the other statistical approaches and the individual deep learning models due to the high feature representation opportunities in the numerous stacked neural networks.

Alyahyan et al. [11] consider multiple attributes when trying to predict the result of a college admission decisions. These attributes are broadly classified into categories such as, prior-academic achievement, student demographics, e-learning activity, psychological attributes, and environments. It is noticed that prior academic achievement and student demographics are the top two factors taking all the research papers into consideration. It is also noticed that studies which take into account university data showed better results than studies which only considered pre-university studies or demographics. Presently, the research available is restricted to certain geographical areas such as only America [12] or the Sichuan province [13] or to a particular college such as liberal arts college in California [14].

Machine Learning models are used for training and testing in different ways by researchers. For example, [15] chose to use ensemble learning for an efficient result. In [16] the performance is analyzed through Weka in order to decide which model performs the best based on mean absolute error (MAE) value. While [14], [17], [18] and [19] use multiple machine learning models and pick the model that performs the best. The performance is analyzed through various metrics such as accuracy, precision, F-measure, recall and area under the receiver operator curve.

Research done in the educational spectrum that is not about graduate admissions is also considered because the models are trained on similar data. For instance, in [20] graduate success of students is predicted using undergraduate performance indicators and their aggregates. An all-round analysis of students is made using 81 variables out of 171 student records from a program in Computer Science. Regression models are considered in combination with variable selection and variable aggregation embedded in a double-layered cross-validation loop. Moreover, bootstrapping is employed to identify the importance of explanatory variables. The results of this study show that undergraduate level performance can explain 54% of the variance in graduate-level performance. Linear regression models are employed in combination with different variable-selection techniques. To avoid potentially misleading results caused by methods that are too simplistic, a rigorous data mining methodology is employed that includes cross-validation to avoid overfitting, bootstrapping to assess the stability of variable selection, and statistical testing to estimate differences in performance. Crucially, the results provide a methodological basis for deriving principled guidelines for admissions committees.

Another such paper [18] presents a case study on predicting performance of students at the end of a university degree at an early stage of the degree program, in order to help universities not only to focus more on bright students but also to initially identify students with low academic achievement and find ways to support them. The data of four academic cohorts comprising 347 undergraduate students is mined with different classifiers. Decision trees, rule induction, artificial neural networks, k-nearest neighbor and naive Bayes algorithm are used to provide the best results. The finals results show that Naïve Bayes resulted in an accuracy of 83.65%.

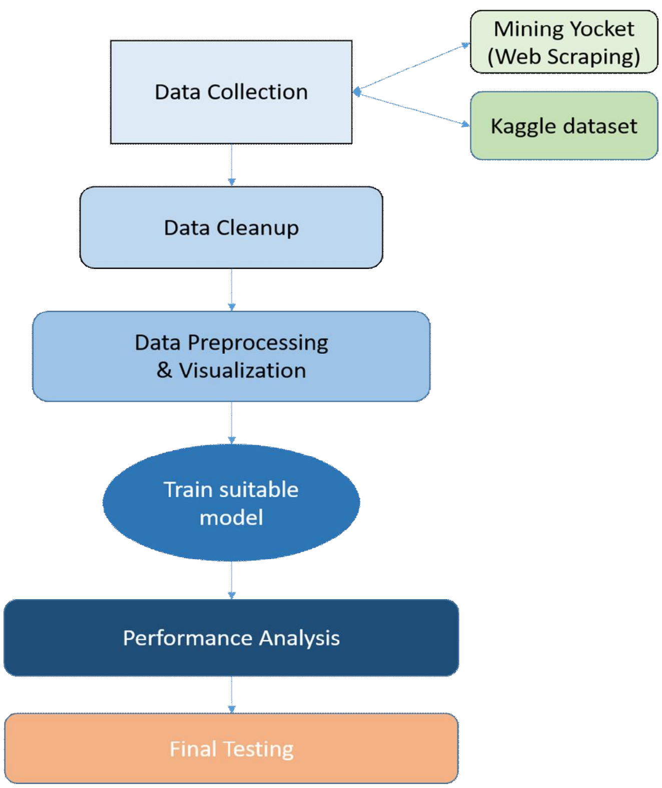

Figure 1: Methodology of the workflow

Use of Deep Learning in Educational data analysis has risen in recent years. Bendangnuksung et al. [21] performed a study where deep learning and ML models are compared. The best accuracy of 84.3% is achieved by deep learning. With a first of its kind Deep Neural Network for measuring students' performance, it is noticed that the Deep Neural Network model performs better than ML models. The added benefit is that lesser amount of data with deep knowledge of the dataset is required. Another similar study is done by Lau et al. [22], the Levenberg–Marquardt algorithm is used as the backpropagation training rule. On the other hand, [23] implement backpropagation with multi-layer feedforward network to predict graduation success. A total of 5100 student samples are compiled for training and testing. The classification accuracy is more than 95% for the best performance achieved. It is mentioned that the same model can be used to predict admission decisions.

3. METHODOLOGY

The methodology followed during the entire workflow is depicted in Figure 1. It consists of five major phases - Data Collection and Cleaning, Data Processing and Visualization, Model Training, Performance Analysis and Testing. The first three stages is discussed in detail in the subsequent subsections, and the Performance analysis and Testing is discussed in Section 4.

A. Data Collection and Training

The dataset was collected by mining 'Yocket' using Selenium WebDriver. First, source URLs from where data was to be mined was manually populated into a csv file. The first column held the URLs of the page that belonged to the admits category and the second column held the rejects. Initially, the 'admits' column was scraped and stored in a csv file. Then, the URLs in the second column was scraped to collect the 'rejects'. This 'rejects' data was also stored in a separate csv file. Lastly, both these csv files were combined.

After the union of the datasets, it became imperative to clean the data to fit our use case and handle the missing values. Thus, the datasets were cleaned to remove unnecessary columns such as Student name, University name, GRE, Eng_Test, Undergrad, and is_work_ex. The Figure 2 represents sample columns in the dataset before cleaning. Furthermore, the missing values in the dataset corresponding to the 'NA' values were imputed with '0'. This might not be the optimal method for imputing the missing values; however, it was done to retain a sufficient number of records for the training phase. Even a mean of the other scores could have been considered but this would falsely represent the actual values. Moreover, the categorical data is encoded to ensure compatibility with all the models and remove any bias during analysis. The values of years '2020' and '2021' is changed to '1' and all the other years to '0' to mark the COVID-19 period. This year-wise marking was done to facilitate the analysis of the effect of COVID-19 on the admissions. Figure 3 is the dataset after merging and cleaning. The attribute wise data distribution is depicted by Figure 4.

	student_name	University_name	year_applying	Status	GRE	GRE_SCORE	Eng_test	Test_score	Undergraduation	Undergraduation_score	is_work_ex	work_ex
0	Pooja Dattatri	sylvania Computer & I	2021	Admit	GRE	321	TOEFL	108	UNDERGRAD	8.91	WORK EX	18
1	dict Florance Aroc	sylvania Computer & I	2021	Admit	GRE	0	ENG TEST	0	UNDERGRAD	9.49	WORK EX	0
2	manni arora	sylvania Computer & I	2020	Admit	GRE	320	TOEFL	110	UNDERGRAD	9.34	WORK EX	24
3	Priti Goyal	sylvania Computer & I	2020	Admit	GRE	324	TOEFL	115	UNDERGRAD	9.35	WORK EX	12
4	PoojaDattatri	sylvania Computer & I	2021	Admit	GRE	321	TOEFL	108	UNDERGRAD	8.91	WORK EX	18
5	benedictflorance	sylvania Computer & I	2021	Admit	GRE	0	ENG TEST	0	UNDERGRAD	9.49	WORK EX	0
6	manniarora	sylvania Computer & I	2020	Admit	GRE	320	TOEFL	110	UNDERGRAD	9.34	WORK EX	24
7	Priti11234578	sylvania Computer & I	2020	Admit	GRE	324	TOEFL	115	UNDERGRAD	9.35	WORK EX	12
8	bhawikaagarwal	sylvania Computer & I	2020	Admit	GRE	327	TOEFL	109	UNDERGRAD	9.38	WORK EX	0

Figure 2: Sample rows of the dataset before being cleaned

B. Data Preprocessing and Visualization

Once the datasets have been combined the next step would be to visualize the data. Histograms, Distplots, heatmaps can be plotted to understand the data and the individual variables. After the data cleaning process, most of the noise and unnecessary data has already been filtered out. The data is now present in a tabular format consisting of student records as the rows and the attributes as the columns. Each column represents a certain characteristic of the student application like CGPA, GRE Score, TOEFL/IELTS Score (Test_Score), Work Experience, Year applying and Decision on Application (Status). Every column has a unique type of information that have various ranges and datatypes. Table 1 provides attribute level information for each attribute.

Table 1: Attribute Information of the Dataset

Attributes	Min. Value	Max. Value	Datatype
CGPA	0	10	Float
GRE Score	260	340	Integer
TOEFL/IELTS Score	0	120	Integer
Work Experience	0	-	Float
Year Applying	0 (Before 2020)	1 (After 2019)	Integer
Decision	False	True	Boolean

All the attributes other than the Decision itself would potentially serve as the input features and the Decision serves as the ground truth for any prediction system. All the input features are numerical type data (either Integer or a Float). This fact reinforces the selection of Min-Max Scaling as the standardization technique. Upon standardization, the range of the data is compressed to the range [0, 1]. This makes it easier for the model to process the information and remove the range bias. The output or the ground truth is converted to a binary {0, 1} format. Various graphs and charts can be prepared owing to the numerical nature of the data. These charts can help exaggerate and highlight the major correlations and preferences of the universities across the globe. Time Series plots can help get an insight on the relation of the admissions based on the time period. This time period analysis is essential for finding and studying the correlation between the COVID-19 pandemic and the impact on the graduate admissions. The major portion of the time period that needs to be considered for the analysis of the impact is the time period from March 2020 to Present.

	year_applying	Status	GRE_SCORE	Test_score	Undergraduation_score	work_ex
0	1	1	321	108	8.91	18
1	1	1	0	0	9.49	0
2	1	1	320	110	9.34	24
3	1	1	324	115	9.35	12
4	1	1	321	108	8.91	18
5	1	1	0	0	9.49	0
6	1	1	320	110	9.34	24
7	1	1	324	115	9.35	12
8	1	1	327	109	9.38	0

Figure 3: Sample rows of the dataset after being cleaned

C. Model Training

The processed data now obtained is used to create prediction models that can aid in predicting whether a prospective candidate gets accepted for the graduate program. This prediction problem is essentially a binary type prediction. A binary prediction is generally regarded as a single class (or in some cases double class) classification. A variety of classification algorithms exist in the domain of Statistics and Machine Learning. We implement a few classification algorithms and use them for comparison. Some implemented algorithms include:

- Random Forest Classification
- Gaussian Naive Bayes Classifier
- Logistic Regression
- Decision Tree Classifier
- Artificial Neural Networks

Random Forest Classification is an ensemble learning method which uses multiple Decision trees with varying parameters or data batches to obtain an optimal classification technique. The ensemble component of the algorithm makes it possible for the model to cover most of the shortcomings of the contributing models making it robust during real time predictions. Gaussian Naive Bayes Classifiers are widely used classifiers to obtain a clear decision boundary for numerical data classification. Since, the data that we are working with consists of numerical data, this makes it a viable option for prediction of graduate admissions data. Logistic Regressions inherently is designed for binary prediction system. Although the system is generally not complex enough to handle high dimensional and subjective data, it serves as a definite benchmark for comparison with other models.

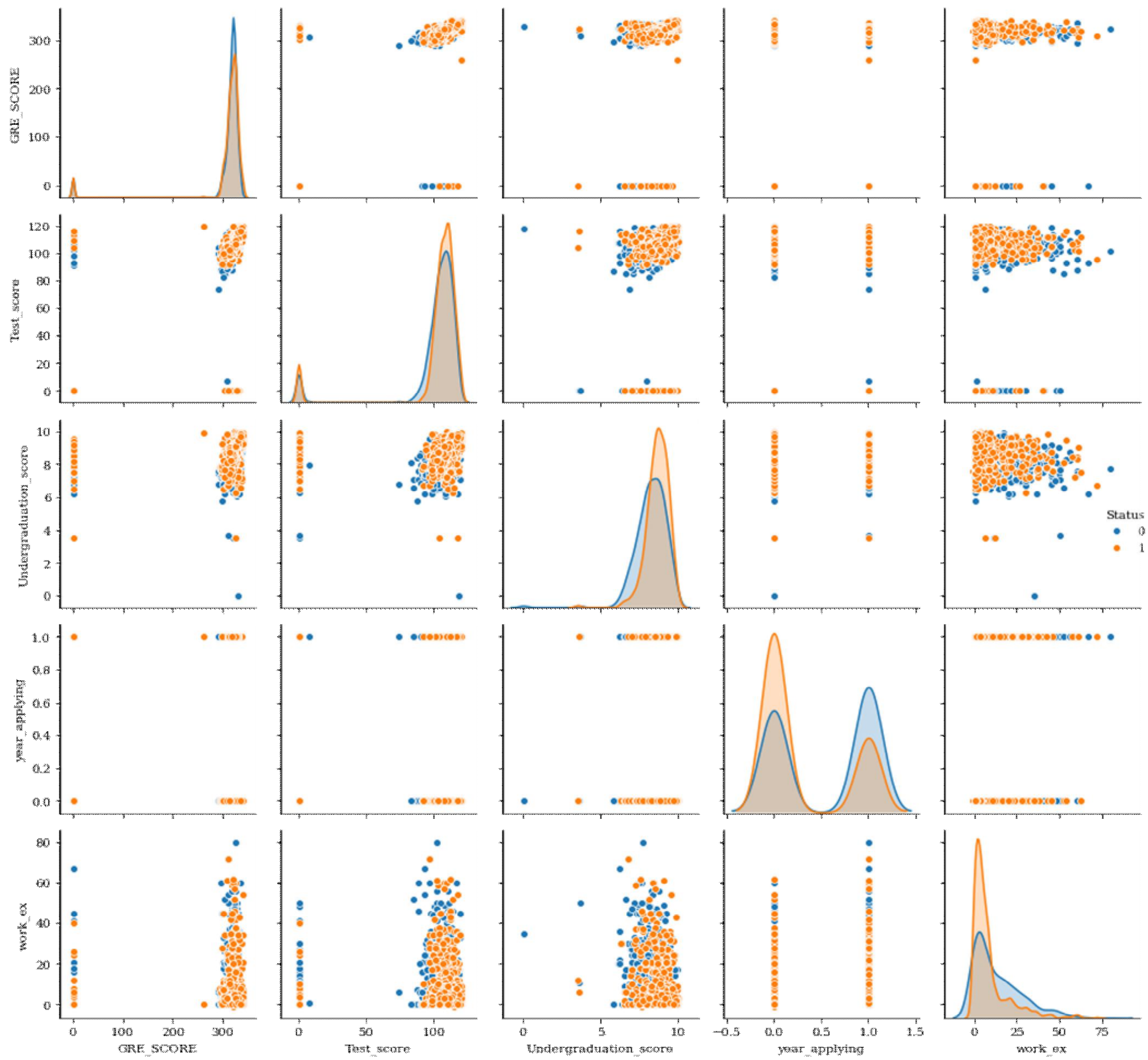

Figure 4: Attribute-wise distribution of data

An Artificial Neural Network is more recent and regarded as one of the most efficient methods of prediction providing near-human accuracy at complex tasks. The ANN in this scenario is trained with a binary cross entropy loss for the forward propagation and coupled with the Adam optimizer for the weight updates in the weight update phase after back propagation. The ANN for binary cross entropy loss generally uses a sigmoid activation function yielding a probability score as the prediction of a positive output. The decision boundary is generally regarded at the 50% threshold, but a better threshold can be obtained from the ROC curve. The ANN starts with an Input layer having the same shape as the input data and them is followed by a series of 2 or 3 Dense layers to ensure proper extraction of features and prevent information bottleneck.

4. RESULTS AND DISCUSSION

This section deals with the experimental results and its corresponding analysis. The training performance subsection states and discusses the performances of the models used and its comparison. The following subsection provides detailed analysis on the effect of COVID19 on the admissions for Graduate Program. The scraping of the 'Yocket' website has been done using a 'Selenium WebDriver' web framework and the 'BeautifulSoup' library. The Random Forest, Gaussian Naive Bayes, Logistic Regression and Decision Tree models have been trained and tested using the 'sklearn' library. The neural network has been defined using 'Tensorflow' and trained on an 'Nvidia RTX 2070' GPU.

A. Training Performance

A The dataset was split for training and testing using the 'test_train_split' function with a test size of 0.05. The maximum depth of both the Decision Tree and Random Forest model was set to 5 for best performance. While the training performance of all the models was measured through various metrics such as accuracy, precision, recall, f1-score and support. The classification report was generated for all the models and the values given are shown in Table 2. The accuracies of each model have also been visualized using a horizontal bar graph in Figure 5. The Neural Networks model resulted in the highest accuracy of 79.03%. The second highest accuracy of 75.80% was achieved by the Random Forest model. The lowest accuracy was recorded by the Gaussian Naive Bayes model, which was 58.06%. The highest precision was reported by the Decision Tree model whereas the lowest was for Gaussian Naive Bayes model. Even for recall, Decision Trees had the highest but Neural Networks had a recall that was lower by just 0.01. But Neural Networks had the highest f1-score and support when compared to all the models. Gaussian Naive Bayes also had the lowest recall and f1-score.

Table 2: Classification report of the models based on different metrics.

Model	Class	Precision	Recall	F1-score	Support
Decision Tree	0.0	0.83	0.61	0.70	31
	1.0	0.69	0.87	0.77	31
Gaussian Naïve Bayes	0.0	0.63	0.39	0.48	31
	1.0	0.56	0.77	0.65	31
Logistic Regression	0.0	0.73	0.52	0.60	31
	1.0	0.62	0.81	0.70	31
Neural Net	0.0	0.80	0.86	0.83	37
	1.0	0.77	0.68	0.72	25
Random Forest	0.0	0.79	0.71	0.75	31
	1.0	0.74	0.81	0.77	31

The neural network is trained for 100 epochs under a variable learning rate to obtain optimal performance. The learning rate is scheduled to half every 45 epochs. This lowering of learning rate provides a controlled gradient descent and limits the overfitting or underfitting in the performance of the neural network. Furthermore, it becomes essential to carefully tune these hyper-parameters in case of less data to obtain the best possible result from the model. After 100 epochs the model settles at about 79.03% accuracy and 0.19 binary crossentropy loss. However, upon analysis of the test performances over all the 100 epochs it can be observed that there is an initial spike in the accuracy of the model. Upon careful examination it is found that the model did not perform on par on other samples of the dataset. This explanation is bolstered by the high loss at this point in training. The entire trend for all the 100 epochs of training is shown in Figure 6.

Figure 7 presents the Receiver Operator Characteristic (ROC) Curve of the trained neural network. The step nature of the curve for the neural network is due to the lack of large amount of data. However, the curvature of the curve tending to peak towards the (0, 1) point can be observed. This coincides with the expectation of the nature of the curve, signifying proper training and performance. The area under the ROC curve (AUROC) score is about 0.877 which is high enough to solidify the existence of a boundary condition for the binary classification. This increases the reliability of prediction of the neural network model in deployment. The fully trained neural network yields an accuracy score of 79.03% which outperforms all the other tested models by an appreciable margin.

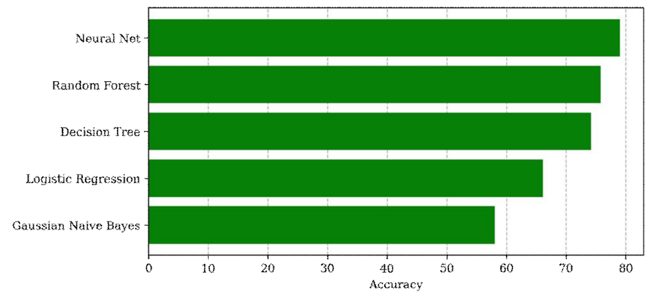

Figure 5: Test accuracy measures of the models

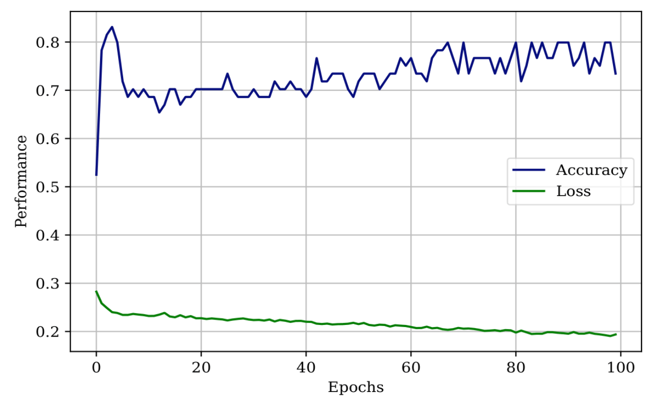

Figure 6: Performance metrics of the Neural Network

B. Analysis

Each one of the models was given four different sample cases in order to perform a comprehensive analysis. The first sample case was designed to analyze the probability of a student receiving an acceptance from a college. The given data for the scores for two students was the same, except the year they applied. Student 1 would have applied before COVID19 while student 2 would apply after COVID19 (After 2020). The probability for student 2 increased in all the models except in random forest where it marginally decreased. The second sample case was used to test how work experience would affect the probability for a student that applied before COVID19 and a student that applied after COVID19. The work experience was increased for the student applying before COVID19. The probability increased for student 1 for Decision Tree and Gaussian Naive Bayes and decreased in the other three models. The third sample case was given the same values from the second example. But for student 2 the year applying was changed to 1, which means both students would be applying after COVID19. The probability increased for student 1 for Decision Tree, Gaussian Naive Bayes and Logistic Regression whereas it decreased in the other two models. The third sample case was given the same values from the second example. But for student 2 the GRE Score was increased. The probability increased for student 1 for Decision Tree while it decreased in the other four models. Based on the resulting probabilities for the four examples provided, it can be stated that a student applying post COVID19 has a higher chance of getting an admit from a college even if they have similar scores as someone applying before COVID19. Although it is clear that colleges are more reliant on GRE scores than before. Even work experience during COVID19 is a major factor that has been taken into consideration. If a student has work during the pandemic their chances of receiving an admission is higher. Similarly, other attributes can also be tested in different ways to analyze how they affect the admissions decision.

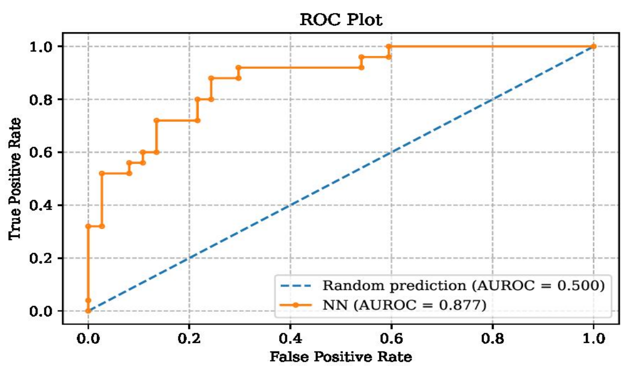

Figure 7: ROC curve for the predictions of the neural network

5. CONCLUSION AND FUTURE SCOPE

In this work, Graduate admissions data is scraped to build a dataset. After the cleaning and preprocessing of this dataset, multiple prediction models are built. These models are trained with an extra parameter signifying whether the record belongs to the COVID-19 pandemic time period. The models used for predictions include Decision Tree, Gaussian Naive Bayes, Logistic Regression, Artificial Neural Networks, and Random Forest Classification. The accuracy scores of the trained models are 74.19%, 58.06%, 66.12%, 79.03%, and 75.80% respectively. The Artificial Neural Network proves to be the best performing model among these tested models. These trained models are then further used to predict the probability of admit in various scenarios. It can be observed from the majority of predictions that probability of a student getting accepted is higher post COVID-19 pandemic. However certain attributes like the work experience was given higher priority before the pandemic. This behavior conveys that the criteria for admits have been eased due to the pandemic situation. Furthermore, the likely cause of certain behavioral traits of the data can be attributed to the lower number of applicants in the duration of the pandemic.

The dataset has been scraped and cleaned thoroughly. However, the lack of public access data restricts the amount of available data. The size of the dataset can be further increased to facilitate the use of more complex models in future implementation. The large dataset size can provide more flexibility during the data cleaning and preprocessing stages. Removal of outliers, intelligent imputation of missing values and higher degree of interference generalization are a few improvements that could potentially increase the reliability of the models both for analysis and deployment. Additionally, a portal implementing these models and analysis can be developed for public use. This portal can provide services to institutes or students to ease the Graduate Admissions process.

Table 3: Probability of students getting admitted into college based on the prediction of different models

Attributes	Example 1 (Same Scores pre-covid and post-covid)		Example 2 (More Work-Experience pre-covid)		Example 3 (More Work-Experience during covid)		Example 4 (Higher GRE Score pre-covid)	
	Student 1	Student 2	Student 1	Student 2	Student 1	Student 2	Student 1	Student 2
Year Applying	1	0	1	0	1	1	1	0
GRE Score	305	305	317	317	317	317	317	325
TOEFL Score	115	115	103	103	103	103	103	103
CGPA	9.23	9.23	8	8	8	8	8	8
Work Exp.	2	2	0	10	0	10	0	0
Decision Tree Probability	50.56%	**84.00%**	**50.56%**	36.79%	**50.56%**	36.79%	**50.56%**	36.79%
Gaussian Naïve Bayes Probability	60.41%	**60.83%**	**64.73%**	61.88%	**64.73%**	61.47%	64.73%	**65.32%**
Logistic Regression Probability	54.81%	**68.93%**	49.18%	**59.07%**	**49.18%**	44.10%	49.18%	**64.15%**
Neural Net Probability	43.02%	**64.44%**	32.65%	**47.14%**	32.65%	**35.24%**	32.65%	**43.17%**
Random Forest	**54.46%**	52.08%	49.79%	**52.08%**	49.79%	**54.46%**	49.79%	**65.84%**

REFERENCES

1. Austin Waters and Risto Miikkulainen. **Grade: Machine learning support for graduate admissions.** Ai Magazine, 35(1):64–64, 2014.
2. Robert L Linn. **Admissions testing: Recommended uses, validity, differential prediction, and coaching.** Applied Measurement in Education, 3(4):297–318, 1990.
3. Mohan S Acharya, Asfia Armaan, and Aneeta S Antony. **A comparison of regression models for prediction of graduate admissions.** In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), pages 1–5. IEEE, 2019.
4. Bhavya Ghai. **Analysis & prediction of american graduate admissions process.** 2018.
5. Selvaprabu Jeganathan, Saravanan Parthasarathy, Arun Raj Lakshminarayanan, PM Ashok Kumar, and Md Khurshid Alam Khan. **Predicting the post graduate admissions using classification techniques.** In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), pages 346–350. IEEE, 2021.
6. Inssaf El Guabassi, Zakaria Bousalem, Rim Marah, and Aimad Qazdar. **A recommender system for predicting students' admission to a graduate program using machine learning algorithms.** 2021.
7. Navoneel Chakrabarty, Siddhartha Chowdhury, and Srinibas Rana. **A statistical approach to graduate admissions' chance prediction.** In Innovations in Computer Science and Engineering, pages333–340. Springer, 2020.
8. Md Omaer Faruq Goni, Abdul Matin, Tonmoy Hasan, Md Abu Ismail Siddique, Oishi Jyoti, and Fahim MD Sifnatul Hasnain. **Graduate admission chance prediction using deep neural network.** In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pages 259–262. IEEE, 2020.
9. P Nandal. **Deep learning in diverse computing and network applications student admission predictor using deep learning.** In Proceedings of the International Conference on Innovative Computing & Communications (ICICC), 2020.
10. Sashank Sridhar, Siddartha Mootha, and Santosh Kolagati. **A university admission prediction system using stacked ensemble learning.** In 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), pages 162–167. IEEE, 2020.
11. Eyman Alyahyan and Dilek Dustegor. **Predicting academic success in higher education literature review and best practices.** International Journal of Educational Technology in Higher Education, 17,012020.
12. Peta R. Bindu S. Swaroop S. Chithra A., Malepati C. **Prediction for university admission using machine learning.** International Journal of Engineering Research and Engineering, V8, 2020.
13. Zhenru Wang and Yijie Shi. **Prediction of the admission lines of college entrance examination based

on machine learning. 2016 2nd IEEE International Conference on Computer and Communications (ICCC), pages 332–335, 2016.

14. Dinesh B. Vaghela and Priyanka Sharma. **Students' admission prediction using grbst with distributed data mining.** 2015.
15. Kanadpriya Basu, Treena S. Basu, R. E. Buckmire, and Nishu Lal. **Predictive models of student college commitment decisions using machine learning.** Data, 4:65, 2019.
16. Sara Aljasmi, Ali Nassif, Ismail Shahin, and Ashraf Elnagar. **Graduate admission prediction using machine learning.** 14, 12 2020.
17. Himanshu Sonawane. **Student Admission Predictor.** PhD thesis, Dublin, National College of Ireland,2017.
18. Raheela Asif, Agathe Merceron, and Mahmood K Pathan. **Predicting student academic performance at degree level: a case study.** International Journal of Intelligent Systems and Applications, 7(1):49–61, 2014.
19. Kamble A. Shinde S. Jadhav S., Patil D. **College recommendation system using machine learning.** International Journal of Engineering Research and Engineering, 2020.
20. Judith Zimmermann, Kay Brodersen, Hans Heinimann, and Joachim Buhmann. **A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance.** 12 2018.
21. Bendangnuksung and D. Venkatesa Prabu. **Students' performance prediction using deep neural network.** 2018.
22. E. Lau, L. Sun, and Qingping Yang. **Modelling, prediction and classification of student academic performance using artificial neural networks.** SN Applied Sciences, 1,082019.
23. Gene Lesinski, Steven Corns, and C. H. Dagli. **Application of an artificial neural network to predict graduation success at the united states military academy.** Procedia Computer Science, 95:375–382, 122016.