

# PROCEEDINGS

## 2020 IEEE 9<sup>th</sup> INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS AND NETWORK TECHNOLOGIES

**Edited by**

*G.S. Tomar, Chairman, IEEE MP Sub Section, India*

**Organized by**

*IEEE MP Sub Section and Hosted by  
SRCEM, Banmore, near Gwalior, India*



**Financial Sponsor: IETE Gwalior, India**



**Technically Co-Sponsored by**



**IEEE Bombay Section**



**Machine Intelligence Research Labs, Gwalior**

# **2020 9<sup>th</sup> IEEE International Conference on Communication Systems and Network Technologies**

## **CSNT 2020 Table of Contents**

<i>Welcome from IEEE CSNT 2020 General Chairs.....</i>	<i>i</i>
<i>Welcome from IEEE CSNT 2020 Program Chairs.....</i>	<i>iii</i>
<i>Conference Committees.....</i>	<i>iv</i>

---

### **Track-A: Microwave Components, Antenna and Propagation**

A Multiband Antenna for WLAN and WiMAX Wireless Applications .....	01
<i>Rishi Parashar; Yogesh Bhomia, Devendra Soni, Ritesh Kumar Saraswat</i>	
Microwave Imaging Based Automatic Crack Detection System Using Machine Learning For Columns .....	05
<i>Prashanth Kannadaguli, Vidya Bhat</i>	
Rectangular Ring Monopole Antenna using Defected Ground Structure for Television White Space Application .....	09
<i>Sonali Jayant Shedge; Vijaya N. Kamble</i>	
Polarization Independent Super Thin Metamaterial Microwave Broadband Absorber for X-Band Application .....	13
<i>Gaurav Chaitanya; Ankit Chandachoriya</i>	

### **Track-B: AI, Fuzzy and Machine Learning**

Demystifying and Anticipating Graduate School Admissions using Machine Learning Algorithms. 19	
<i>Mohd Aijaz Khan; Manish Dixit; Aaradhya Dixit</i>	
Deep Learning prospective for massive MIMO: Challenges and Future prospects .....	26
<i>Vandana Bhatia; Malay Ranjan Tripathy; Priya Ranjan</i>	
Case for Dynamic Parallelisation using Learning Techniques .....	32
<i>Karthik Gurunathan; Kaustubh Kartikey; Sudarshan TSB; Divyaprabha K.N.</i>	
Implementation of Machine Learning Technique for Identification of Yoga Poses .....	40
<i>Yash Agrawal; Yash Shah; Abhishek Sharma</i>	
Transfer Learning with L2 Norm Regularization for classifying static Two Hand Hindi Sign Language Gestures .....	44
<i>Mohita Jaiswal; Vaidehi Sharma; Abhishek Sharma; Raghuvir Tomar</i>	
Hand Written Digit Recognition using Machine Learning .....	49
<i>Rohan Sethi; Ila Kaushik</i>	

Accelerated Computer Vision Inference with AI on the Edge .....	55
<i>Varnit Mittal; Bharat Bhushan</i>	
IoT Based CNC Machine Condition Monitoring System Using Machine Learning Techniques ...	61
<i>Mohan krishna K; Prashanth Kannadaguli</i>	
Comparative analysis of Human Face Recognition by Traditional Methods and Deep Learning in Real-time Environment .....	66
<i>Ruchi Jayaswal, Manish Dixit</i>	
<b>Track C: Mobile Computing, WSN, IOT and next Generation Networks</b>	
Precedence & Issues of IoT based on Edge Computing .....	72
<i>Sukriti Goyal; Nikhil Sharma; Ila Kaushik; Bharat Bhushan; Abhijeet Kumar</i>	
IOT Based Wireless Sensor Network for Air Pollution Monitoring .....	78
<i>Ajay Chaturvedi; Laxmi Shrivastava</i>	
Network Path Capability Identification and Performance analysis of Mobile Ad hoc Network .....	82
<i>Mohan Patsariya; Anand Rajavat</i>	
Video Transmission Over Next Generation Emergency Services .....	88
<i>Abubkr Elmnsi; Fahad Mira</i>	
Spectrum Sensing Techniques for Cognitive Radio: A Re examination .....	93
<i>Ashish Bagwari; Sonal Tuteja; Ashraf Samarah</i>	
Effect of WSN nearest neighbours on convergence rate of periodic gossip algorithms .....	96
<i>Sateeshkrishna Dhuli; Yatindra Nath Singh; Priya Ranjan</i>	
Importunity & Evolution of IoT for 5G .....	102
<i>Anas Ahmad; Bharat Bhushan; Nikhil Sharma; Ila Kaushik; Saurabh Arora</i>	
Evolution of 5G Wireless Network in IoT .....	108
<i>Saurabh Arora; Nikhil Sharma; Bharat Bhushan; Ila Kaushik; Anas Ahmad</i>	
IoT: A Boon & Threat to the Mankind .....	114
<i>Apeksha Rustagi; Chinkit Manchanda; Nikhil Sharma</i>	
A Survey on IOT enabled cloud platforms .....	120
<i>Ranjana Sikarwar, Pradeep Yadav; Aditya Dubey</i>	
IoT and BLE Beacons: Demand, Challenges, Requirements, and Research Opportunities-Planning-Strategy .....	125
<i>Sagar Damoday Padiya, Vijay S. Gulhane</i>	
Intelligent Monitoring systems for transportation of perishable products based on Internet of Things (IOT).....	130
<i>Mohammad Hossein Ahmadzadegan; Hamidreza Ghorbani; Anna Stahlbrost</i>	
Improved Performance Using Fuzzy Possibilistic C-Means Clustering Algorithm in Wireless Sensor Network .....	134
<i>Shweta Kushwah, Kuldeep Singh Jadon</i>	
A Survey on IOT Based Smart Waste Bin .....	140
<i>Dharmendra Kumar Tripathi; Sandeep Dubey; Sandeep Kumar Agrawal</i>	
State of the art of Mobile Banking Services and Future Prospects in Developing Countries.....	145
<i>Mafizur Rahman; Malika Tazim; Sumita Das; Linta Islam</i>	
Emotion Classification on Twitter Data Using Word Embedding and Lexicon based approach.....	150
<i>R Jeberson Retna Raj; Prasanjeet Das; Prabat Sahu</i>	

## **Track D: Hardware Design, Data Mining and systems**

PWM Techniques to Power converters of the Wind Energy Conversion System .....	155
<i>M Satyendra Kumar, K Latha Shenoy; G.B. Praveen</i>	
COMPASS: IPS-based Navigation System for Visually Impaired Students .....	161
<i>Munira AlZamil; Reema Albugmi; Shatha AlOtaibi; Ghadeer AlAnazi; Loay Alzubaidi; Abul Bashar</i>	
Implementation and Extension of Bit Manipulation Instruction on RISC-V Architecture using FPGA .....	167
<i>Vineet Jain; Abhishek Sharma; Eduardo A Bezerra</i>	
Producing Energy Using Blind Man Stick .....	173
<i>Depender Kumar Soni; Nikhil Sharma; Ila Kaushik; Bharat Bhushan</i>	
Optimization of Cost: Storage over Cloud Versus on Premises Storage.....	179
<i>Rajesh Sen; Akrafi Sharma</i>	
Educational Data Mining Methods: A Survey .....	182
<i>Abdul Aleem; Manoj Madhava Gore</i>	

## **Track E: Cryptography, Blockchain and Security Algorithms**

A Survey on Hardware Implementation of Cryptographic Algorithms Using Field Programmable Gate Array.....	189
<i>Keshav Kumar; K. R. Ramkumar; Amanpreet Kaur; Somanshu Choudhary</i>	
Secured Multi-Tier Mutual Authentication Protocol for Secure IoT System.....	195
<i>Rupali S. Vairagade; Bramhananda S.H.</i>	
Improved AES with a Privacy Database Structure for IoT Nodes.....	201
<i>Joseph Henry Anajemba; Celestine Iwendi; Mohit Mittal; Tang Yue</i>	
Design and Development of trust management scheme for internet of things based on the optimization algorithm.....	207
<i>Shilpa V. Shankhpal; Bramhananda S.H.</i>	
High Dimensional Data Processing in Privacy Preserving Data Mining.....	212
<i>Mayur Rath; Anand Rajavat</i>	
IoT based Smart home: Security Aspects and security architecture.....	218
<i>Abhay Kumar Ray; Ashish Bagwari</i>	
Security Issues & Seclusion in Bitcoin System.....	223
<i>Depender Soni; Harbhajan Sharma; Bharat Bhushan; Nikhil Sharma; Ila Kaushik</i>	
Smart Contract Definition for Land Registry in Blockchain.....	230
<i>Archana Sahai; Rajiv Pandey</i>	
Neoteric Security and Privacy Sanctuary Technologies in Smart Cities.....	236
<i>Chinkit Manchanda; Nikhil Sharma; Rajat Rath; Bharat Bhushan; Moksh Grover</i>	
Security Challenges & Controls in Cyber Physical System.....	242
<i>Rajat Rath; Nikhil Sharma; Chinkit Manchanda; Moksh Grover; Bharat Bhushan</i>	
A Comprehensive Survey on various Security Authentication Schemes for Mobile Touch Screen.....	248
<i>Gaurav Bhatt ; Bharat Bhushan</i>	
Working principle, Application areas and Challenges for Blockchain Technology.....	254
<i>Lakshit Madaan; Amit Kumar; Bharat Bhushan</i>	
Blockchain for Cybersecurity: A Comprehensive Survey.....	260
<i>Pranshu Bansal; Rohit Panchal; Sarthak Bassi; Amit Kumar</i>	

A Multi-classifier Framework for Detecting Spam and Fake Spam Messages in Twitter.....	266
<i>R Jeberson Retna Raj; Senduru Srinivasulu, Aldrin Ashutosh</i>	
Security-Centric Investigation of Social Networks and Preventative Behavioral Analysis of Online Activity by the Kuleshov effect.....	271
<i>Mohammad Hossein Ahmadzadegan; Hamidreza Ghorbani; Anna Stahlbrost</i>	
An Efficient Storage in the cloud & Secure EHR Retrieval by using HECC.....	277
<i>Poonam Kumari; Neetesh Kr Gupta</i>	
A Comprehensive Survey on Various Machine Learning Methods used for Intrusion Detection System.....	282
<i>Akshay Gupta, Jitendra Agrawal</i>	
<b>Track F: Image Processing, Signal Processing intelligent systems</b>	
Effective Framework for Underwater Image Enhancement using Multi-Fusion Technique.....	290
<i>Aashi Singh; Khushboo Agarwal</i>	
Side searching and object improving algorithms for Images.....	296
<i>Ranbeer Tyagi, GS Tomar, Laxmi Shrivastava</i>	
Phishing Detection Using Significant Feature Selection.....	302
<i>D.N. Goswami; Manali Shukla; Anshu Chaturvedi</i>	
Analysis of Geographical Change Detection using Satellite Images.....	306
<i>Shubhangi Yerne ; Urmila Shrawankar</i>	
BDonor: A Geo-localised Blood Donor Management System Using Mobile Crowdsourcing.....	313
<i>Hridoy Deb Das; Rakib Ahmed; Nurunnahar Smrity; Linta Islam</i>	
An Analysis of Image Segmentation Methods for Brain Tumour Detection on MRI Images.....	318
<i>Anurag Goswami; Manish Dixit</i>	
Wavelet Based Empirical Approach to Mitigate the Effect of Motion Artifacts from EEG Signal..	323
<i>Shailja Shukla; Vandana Roy; Anand Prakash</i>	
Different Techniques for Identification of a Bone Fracture in Analysis of Medical Image.....	327
<i>Rinisha Bagaria; Sulochna Wadhvani; A.K. Wadhvani</i>	
A Survey on Underwater Images Enhancement Techniques.....	333
<i>Om Kumari Soni, Jamvant Singh Kumre</i>	
<b>Author Index .....</b>	<b>339</b>

# Demystifying and Anticipating Graduate School Admissions using Machine Learning Algorithms

Mohd Aijaj Khan, Manish Dixit  
Madhav Institute of Technology and Science  
Gwalior, India  
[AijajKhan@ieee.org](mailto:AijajKhan@ieee.org), [Manishdixit@ieee.org](mailto:Manishdixit@ieee.org)

Aaradhya Dixit  
Electronics & Communication Engg, Birla Institute of  
Technology, Mesra, India,  
[AaradhyaDixit@ieee.org](mailto:AaradhyaDixit@ieee.org)

**Abstract**—One of the many aspirations of undergraduate students in India is going for further graduate studies. Unfortunately, many students spend months and years of preparation focusing on things that unfortunately won't improve their chances of getting into a good graduate school. This paper evaluates the chances of applicants to get into a particular graduate program using various classification and regression approaches of Machine Learning. Various algorithms have been pitted against each other and also the most important features have been extracted which are useful to get into a graduate school program. Using unsupervised approach, this paper finds various categories of students and pool them together to find if they are perfect fit for admission or not. A novel approach of predicting the chances for admission in graduate school is introduced in this paper.

**Keywords**—Machine Learning, Support Vector Machines, Artificial Neural Networks, Elbow Method, Keras, Python, Scikit, Pandas, Features, Redundancy, Supervised Learning, Unsupervised Learning, Cumulative Grade Point Average (CGPA), Graduate Record Examinations (GRE), Test of English as a Foreign Language (TOEFL), Accuracy, F Score,  $R^2$  Score, Hidden Layers, Confusion Matrix, Epoch, Metric, Regression, Classification.

## I. INTRODUCTION

The complicated procedure of applying to a graduate school in the USA results in a very hectic undergraduate schedule for Indian students. The parallel approach of keeping a high Cumulative Grade Point Average, with good GRE, TOEFL scores and publishing research papers, getting good Letters of Recommendation and making a good Statement of Purpose certainly makes every Indian student busy who wants to further excel in academia. [1] We understand that all the things cumulatively are very important for graduate school admissions. But what are the most important factors? If we can find those most important factors, we can emphasize more on them to increase our chances to get an admit. Given a labeled dataset of 500 students who applied to a graduate program, we will find the machine learning algorithm which will very closely predict the chances of admission. [4] And from these techniques, we will also extract some of the redundant and very important features. This paper also takes an approach to find the relation of the features for

evaluating the chances of admit from a graduate school. We will also convert it into a classification problem and similarly evaluate a confusion matrix with the aid of classification algorithms and the dataset. Several Machine Learning techniques have been used here and comparative analysis on results of every approach has been done to formulate a novel approach to predict the probability of admission. Also, many powerful techniques such as Support Vector Machines and Artificial Neural Networks have also been used to predict the same. [3]

## II. APPROACHES USED

Given a dataset of 500 students, who had applied to the same graduate program earlier, we see the trend of the features, try to manipulate them, remove redundant features and fit them into models for predicting the chances of admission. [4] Labels are also converted into binary format for classification. The models used are as follows. [2][3]

### A. Regression Algorithms

- Linear Regression
- Support Vector Regression [7]
- Ridge Regression
- Bayesian Ridge Regression
- Artificial Neural Networks
- Random Forest Regression [6]
- Ada Boost Regression
- K-Nearest Neighbors Regression
- Decision Tree Regression

### B. Classification Algorithms

- Artificial Neural Networks
- Logistic Regression
- Naïve Bayes Classifier
- Support Vector Machines [7]
- Perceptron
- K-Nearest Neighbors Classifier
- Random Forest Classifier [6]
- Decision Tree Classifier

We will also cluster our data into various clusters by K-Means Clustering for distinguishing the low-performing applicants from the outstanding ones to help them to prepare more to achieve their goals. We will be using Python as our programming language along with Numpy, Scikit-learn, Keras for Machine Learning, Pandas for data



manipulation and Matplotlib and Seaborn libraries for data visualization and analyzing our learning algorithms. [5]

### III. MACHINE LEARNING AND FEATURE MANIPULATION

Machine Learning is a way to train our machines to learn certain things on their own without programming them explicitly for every task. The machine learns using the learning algorithms, do some task repeatedly and learn from their experience. For Machine Learning problems, first we must analyze our features, plot the input variables against the target and see which features are important and which ones are redundant. We will also see that if there is also any feature that is looking unimportant but is essential for getting our data fit. We will also try to map correlation between different factors that will help to analyze our dataset. [8]. First, we will import our dataset by using Pandas, and we will see the columns and for any missing values. Columns are as follows with the range of their values:

- Serial No. (1-500)
- GRE Score (260-340)
- TOEFL Score (0-120)
- University Rating (1-5)
- SOP (1-5)
- LOR (1-5)
- CGPA (0-10)
- Research (0 or 1)
- Chance of Admit (0 to 1)

Here, “Chance of Admit” is our label (target value) whose value is between 0 and 1 (Probability of getting admitted). Serial No. is a redundant field and we don’t require it, so we will be dropping it from our dataset. CGPA has a mean of 8.57 and its Standard deviation is 0.6. GRE Score has a mean of 316 and its Standard deviation is 11. TOEFL Score has a mean of 107 and its Standard deviation is 6. Which means the scores of the applicants do not differ by much. Also, our Research (value 1 or 0) has a mean of 0.56 which means 56% of applicants have done research work.

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	250.500000	316.472000	107.192000	3.114000	3.374000	3.484000	8.576440	0.560000	0.72174
std	144.481833	11.295148	6.081868	1.143512	0.991004	0.92545	0.604813	0.496884	0.14114
min	1.000000	290.000000	92.000000	1.000000	1.000000	1.000000	6.800000	0.000000	0.34000
25%	125.750000	308.000000	103.000000	2.000000	2.500000	3.000000	8.127500	0.000000	0.63000
50%	250.500000	317.000000	107.000000	3.000000	3.500000	3.500000	8.560000	1.000000	0.72000
75%	375.250000	325.000000	112.000000	4.000000	4.000000	4.000000	9.040000	1.000000	0.82000
max	500.000000	340.000000	120.000000	5.000000	5.000000	5.000000	9.920000	1.000000	0.97000

Fig 1. Information of our dataset including median, standard deviation and mean

What we infer from the data is that, GRE scores are directly proportional to the chance of admit. And so, are TOEFL and CGPA. The Standard Deviation for GRE is 11.29 and TOEFL is 6.08. The Scores of applicants do not differ by much. We will boxplot GRE, TOEFL and CGPA scores to see the distribution,

mean, median of scores and outliers, if any

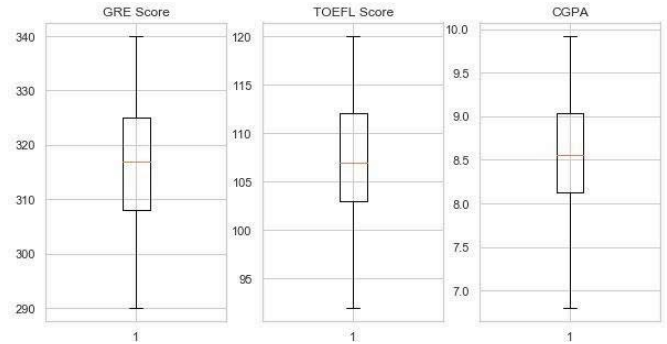


Fig 2. Box Plot to see any abnormalities in the distribution of features.

To find the correlation between different features, we will plot the heatmap of correlation matrix where darker shades denote a large correlation and lighter ones means they are less dependent on each other. The most important column here is to see the “Chance of Admit” to find which factors weigh more for admission.

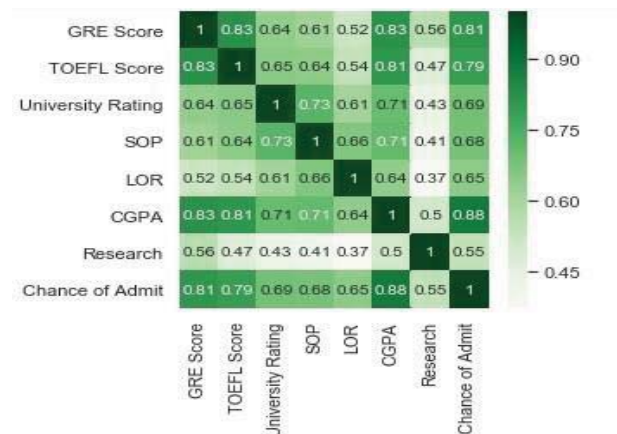


Fig 3. Correlation of different features of dataset using heatmap

We find here that chance of admit has a high correlation with CGPA, GRE and TOEFL scores. Taking the values, we get as expected:

GRE Score	0.810351
TOEFL Score	0.792228
University Rating	0.690132
SOP	0.684137
LOR	0.645365
CGPA	0.882413
Research	0.545871
Chance of Admit	1.000000

Now, we will plot the curve between CGPA and

Chance of Admit and we will find that the relation between them is linear

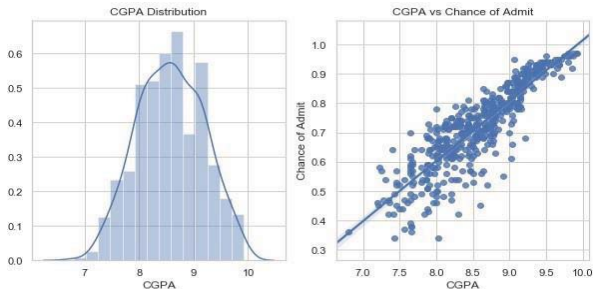


Fig 4. Distribution of CGPA and the Linear Relation of CGPA vs Chance of Admit

Also, the relation between GRE Scores and target variable is linear and the same holds true for CGPA. Hence, the assumptions are verified. The distribution graphs for GRE scores are also plotted. The same is applicable for TOEFL scores according to dataset. [5]

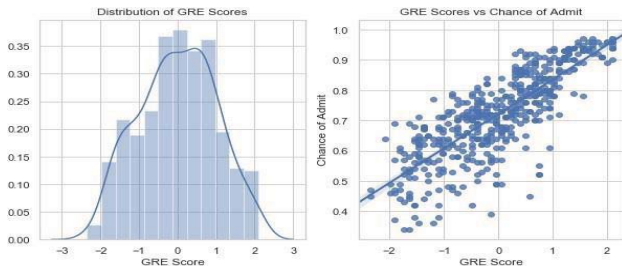


Fig 5. Distribution of GRE scores and the plot of GRE against Chance of Admit (Linear)

We know that Research Experience also plays a very important role in Grad School admissions. Recall from earlier that only 56% of the applicants had research experience.

Total Students: 500

Students having research: 280

Students not having research: 220

Percentage of students having research: 56.0%

The factor plot of Research against Chance of Admit and the distribution graphs are shown in Figure 6.

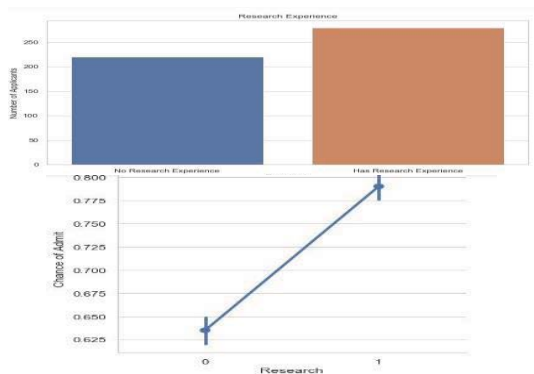


Fig 6. Count of Applicants by Research done and Factor Plot of Research on Admit

And, University rating also plays a very big role. Say, an applicant from IIT having same credentials as compared to an applicant from a Tier 3 university will have a higher chance of getting admission into the Graduate school. For this dataset, the higher the rating, the lower the Tier. Example- University Rating of 5 means Tier 1 university, which is the best of all tiers of universities. The number of students who applied from various tiers of universities are plotted in the next graph shown in Figure 7.

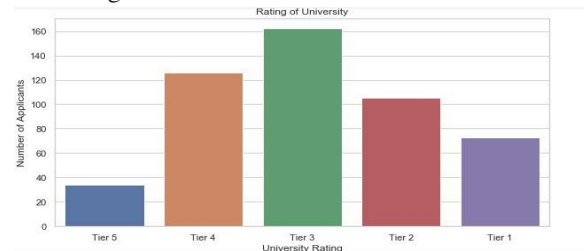


Fig 7. Number of Applicants from different tiers of universities.

The chance of admit vs rating is also given below in Fig 8:

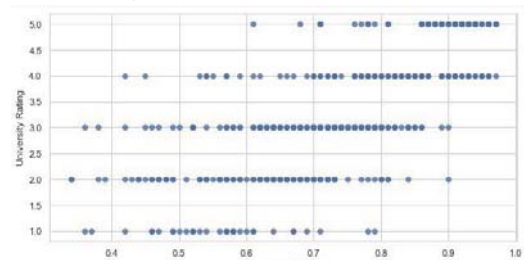


Fig 8. Chance of Admit given University Rating

Letters of Recommendation are being seen redundant here. But, they are related to interaction between professors and students under whose supervision they have performed research work. [5] The students who have done research must have more chances of admission. They are plotted in Fig.9:

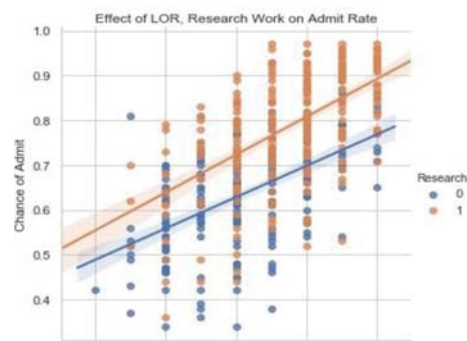


Fig.9: Effect of LOR in conjunction with research work on Chance of Admission



Now we will perform mean normalization of the columns of GRE and TOEFL as they have high ranges. [8] The formula is:

$$X_{i\_new} = (X_i - \text{Mean}) / \text{Standard Deviation} \quad \dots(1)$$

#### IV. APPLYING MACHINE LEARNING REGRESSORS AND OBSERVING RESPONSE

Now we will apply various machine learning regressors and classifiers to our problem and note the responses. First, we will split our dataset into train set and test set by the ratio of 74:26. Here, MSE stands for Mean Squared Error and R Square for  $R^2$  score. [2]

##### A. Linear Regression

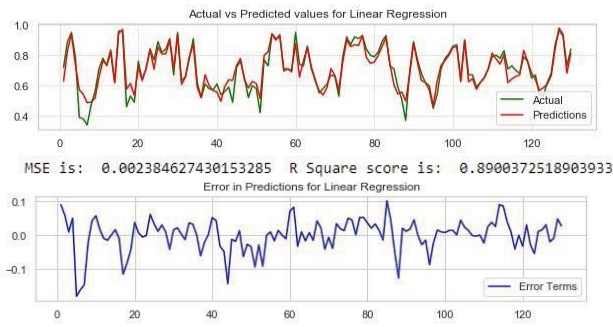


Fig 10. Predictions and Error of Linear Regression

##### B. Support Vector Regression

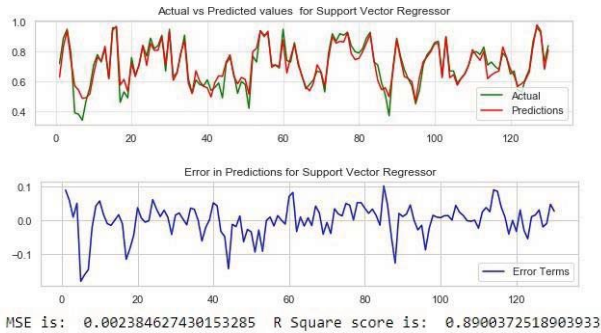


Fig 11. Predictions and Error of Support Vector Regression

##### C. Ridge Regression

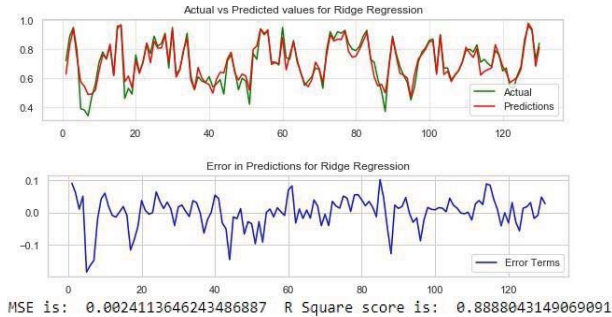


Fig. 12: Predictions and Error of Ridge Regression

##### D. Bayesian Ridge Regression

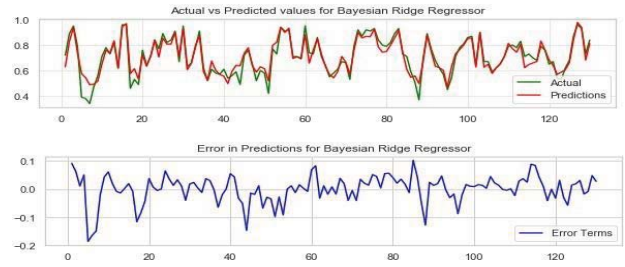


Fig.13: Predictions and Error of Bayesian Ridge Regression

##### E. Artificial Neural Networks

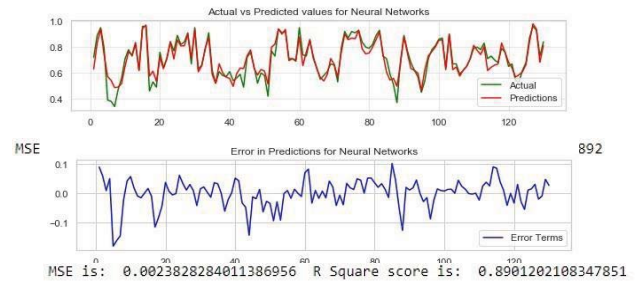


Fig.14: Predictions and Error of Artificial Neural Network

##### F. Random Forests Regression [6]

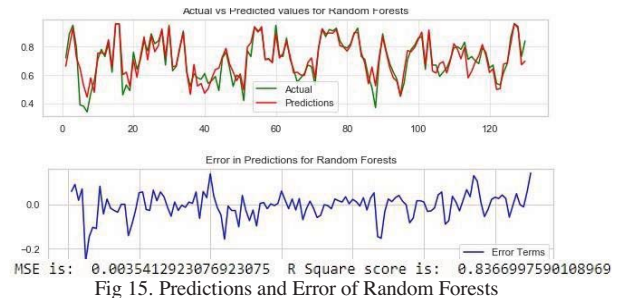


Fig 15. Predictions and Error of Random Forests

##### G. Ada Boost Regression

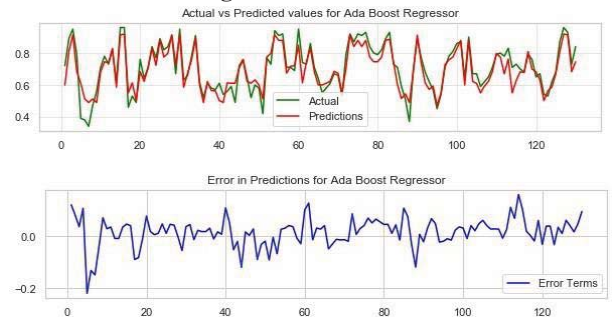


Fig 16. Predictions and Error of Ada Boost Regression

### H. K-Nearest Neighbors

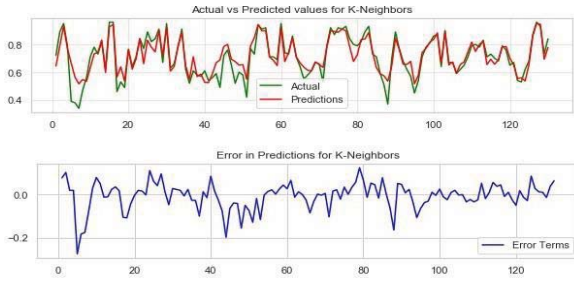


Fig 17. Predictions and Error of K-Nearest Neighbors Regression

### H. Decision Trees

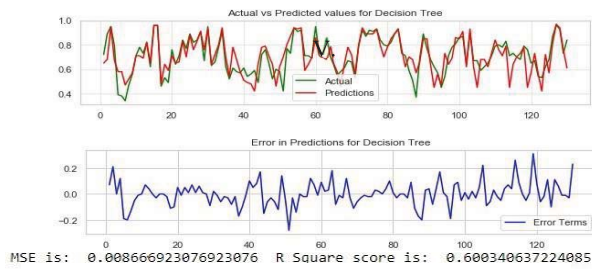


Fig 18. Predictions and Error of Decision Tree Regression

## V. APPLYING MACHINE LEARNING CLASSIFIERS AND OBSERVING RESPONSE

Here, we will first change our data set labels from continuous values to discrete values (0 or 1: Admit or Not) by this algorithm:

$\text{new\_y\_train} = [1 \text{ if } x > 0.8 \text{ else } 0 \text{ for } x \text{ in } y\_train] \dots (2)$

$\text{new\_y\_test} = [1 \text{ if } x > 0.8 \text{ else } 0 \text{ for } x \text{ in } y\_test] \dots (3)$

We selected 0.8 as our threshold here because it was giving the best values of precision and recall for the below algorithms. Plotting heat map of confusion matrix and calculating metrics by the following classification approaches. [5]

#### A. Logistic Regression

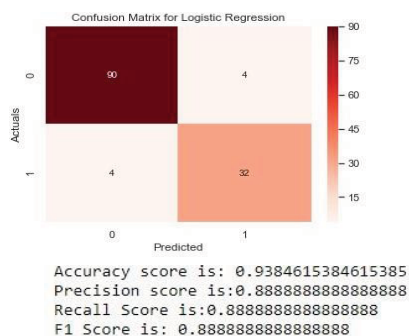


Fig 19. Confusion Matrix and metrics for Logistic Regression

#### B. Support Vector Machine [7]

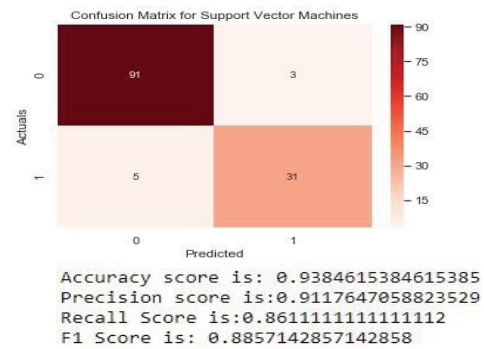


Fig 20. Confusion Matrix and metrics for Support Vector Machines

#### C. Naive Bayes Classifier

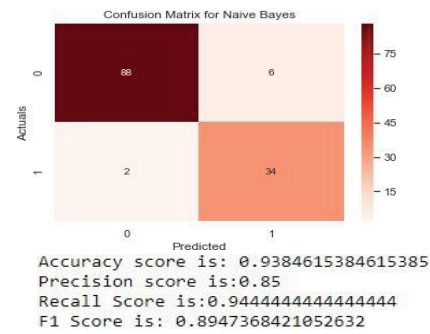


Fig 21. Confusion Matrix and metrics for Naive Bayes Classifier

#### D. Perceptron

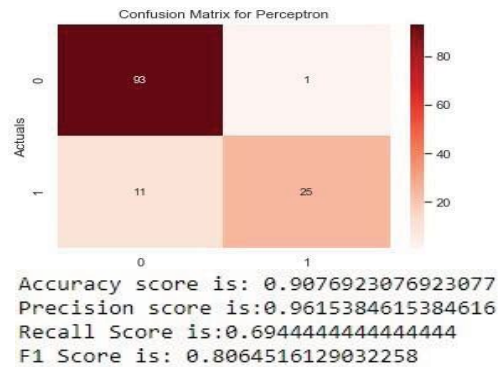


Fig 22. Confusion Matrix and metrics for Perceptron

### E. Decision Tree Classifier

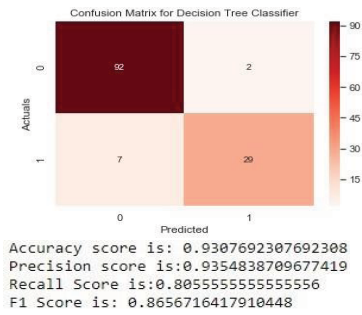


Fig 23. Confusion Matrix and metrics for Decision Tree Classifier

### F. Random Forest Classifier [6]

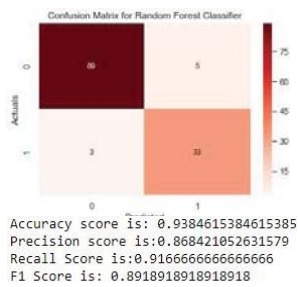


Fig 24. Confusion Matrix and metrics for Random Forest Classifier

### G. K-Nearest Neighbors Classifier

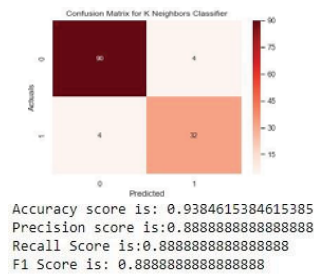


Fig 25. Confusion Matrix and metrics for K-Nearest Neighbors Classifier

### H. Artificial Neural Network

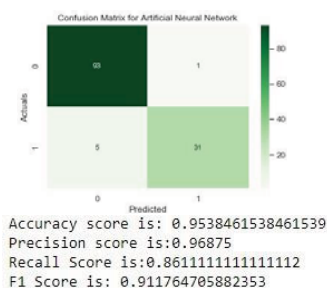


Fig 26. Confusion Matrix and metrics for Artificial Neural Network

## VI. CLUSTERIZING APPLICANTS

Now, we will form clusters of similar applicants and label them if they are a correct fit for admission. [8] Elbow

Method will be used to discover the best number of centroids.

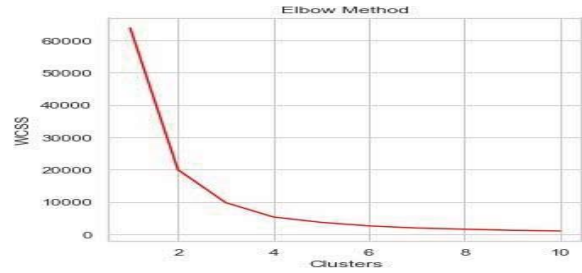


Fig 27. Plot of Least Sum of Squares vs Number of Clusters (Elbow Method)

We will take four clusters according to Elbow method. Plotting and labeling the clusters which are essentially the group of similar applicants pooled together based on CGPA and GRE Score. The labels are also shown in the graph

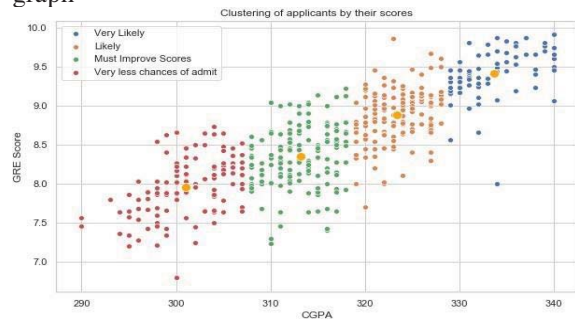


Fig 28. Formation of clusters with centroids and legends

## VII. COMPARISON OF MACHINE LEARNING TECHNIQUES

For finding the best model to fit to this dataset, we compare all the algorithms with their metrics

### A. Regression Algorithms



Fig 29. Performance of Regression Algorithms with respect to  $R^2$  Score and Confidence score

Analysis shows that, our Artificial Neural Network is the best fit here (Accuracy =0.89) and Decision tree overfits the data more according to the confidence score, that is 1.0 [3] [5].

### B. Classification Algorithms

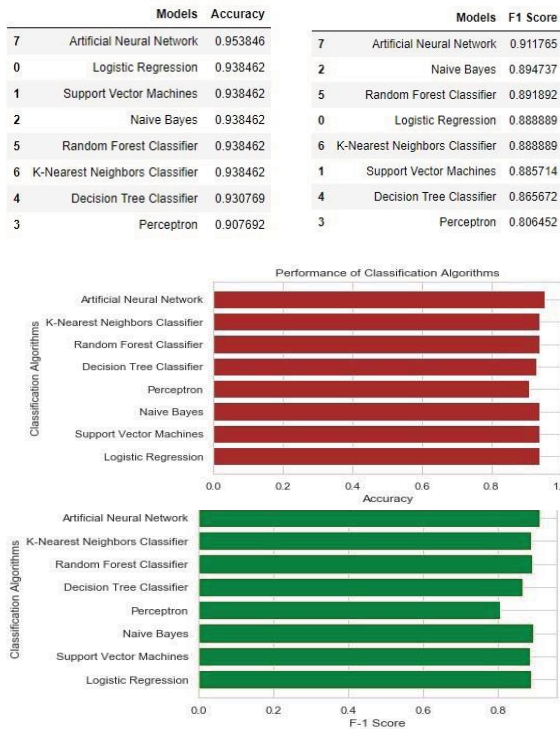


Fig 30. Comparison of Classification Algorithms based on Accuracy and F-Score

## VII. CONCLUSION

Among our classification algorithms, Artificial Neural Network again performs best the based on Accuracy and F-Score (0.95 and 0.91). Even though it does not guarantee to converge to the global minima, it certainly does find a good optimum based on some random initialization of weights [3]. The work has yielded good results and are in accordance with the approach and performance has been satisfactory.

### REFERENCES

- [1] Raghunathan, Karthik. "Demystifying the American graduate admissions process." Web. 2010.
- [2] Gupta, Narender, Aman Sawhney, and Dan Roth. "Will I get in? modeling the graduate admission process for American universities." 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). IEEE, 2016.
- [3] Caruana, Rich, and Alexandru Niculescu-Mizil. "An empirical comparison of supervised learning algorithms." Proceedings of the 23rd international conference on Machine learning. ACM, 2006.

- [4] Mohan Acharya, Asfia Armaan, Aneeta Antony : A Comparison of Regression Models for Prediction of Graduate Admissions, International Conference on Computational Intelligence in Data Science 2019
- [5] Géron, Aurélien. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. "O'Reilly Media, Inc.", 2017.
- [6] Dietterich, Thomas G. "Ensemble methods in machine learning." International workshop on multiple classifier systems. Springer, Berlin, Heidelberg, 2000.
- [7] Schölkopf, Bernhard, Alexander J. Smola, and Francis Bach. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.
- [8] Russell, Stuart J., and Peter Norvig. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited, 2016.