

# A Comparative Study on University Admission Predictions Using Machine Learning Techniques

Prince Golden<sup>1</sup>, Kasturi Mojesh<sup>1</sup>, Lakshmi Madhavi Devarapalli<sup>1</sup>, Pabbidi Naga Suba Reddy<sup>1</sup>, Srigiri Rajesh<sup>1</sup>,  
Ankita Chawla<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, Lovely Professional University,  
Phagwara, Punjab, India

## Article Info

Volume 7, Issue 2

Page Number: 537-548

## Publication Issue :

March-April-2021

## Article History

Accepted : 20 April 2021

Published : 28 April 2021

## ABSTRACT

In this era of Cloud Computing and Machine Learning where every kind of work is getting automated through machine learning techniques running off of cloud servers to complete them more efficiently and quickly, what needs to be addressed is how we are changing our education systems and minimizing the troubles related to our education systems with all the advancements in technology. One of the prominent issues in front of students has always been their graduate admissions and the colleges they should apply to. It has always been difficult to decide as to which university or college should they apply according to their marks obtained during their undergrad as not only it's a tedious and time consuming thing to apply for number of universities at a single time but also expensive. Thus many machine learning solutions have emerged in the recent years to tackle this problem and provide various predictions, estimations and consultancies so that students can easily make their decisions about applying to the universities with higher chances of admission. In this paper, we review the machine learning techniques which are prevalent and provide accurate predictions regarding university admissions. We compare different regression models and machine learning methodologies such as, Random Forest, Linear Regression, Stacked Ensemble Learning, Support Vector Regression, Decision Trees, KNN(K-Nearest Neighbor) etc, used by other authors in their works and try to reach on a conclusion as to which technique will provide better accuracy.

**Keywords** - Linear Regression, Regression Models, Machine Learning Models for prediction, Admission Predictions, Post Graduate studies, Prediction System, Data Mining Techniques.

## I. INTRODUCTION

Undergrad students who wish to go for higher education often start their preparations for their admissions in their dream universities and college

during their final year of their undergrad programme, going through various tests such as GATE, CAT, GRE, TOEFL, IELTS etc. Good scores in these exams with their GPA from their undergraduate programme insures their admissions in their dream colleges.

Students want to pursue their graduation from best universities and colleges with good rankings worldwide. For that many choose to study abroad and thus they take a slew of above mentioned tests that makes it easier for them to apply and get selected for admissions in their preferred institutions. So the graduate program takes immense preparation and it is one of the exhaustive tasks there is to apply and get selected for one's preferred institutions [1]. Hence, to lessen the troubles faced by the students and make their lives easier many efforts has been put through various researchers using ML techniques using various regression models. Here in this paper, we will be reviewing all those works. We will also compare the methods which has been used by other authors to render a model and try to find out which methodology is the most efficient and by how much using the data provided by them. As most of our students try to attain their higher education abroad and most likely from USA[2], we will be using the data set used in the universities of USA such as GRE, GPA, SOP, LOR for the admission predictions which are readily available in the UCLA graduate database for admissions [1]. An efficient university admission predictor will be very helpful as it will narrow down the options available to a student for his/her application to any university as well as give them a certain idea of the chances of their admission in particular university thus eliminating the need to waste time and money on approaching different consultancies and at the same time helps the students to be more informed about the situation [3].

Basically the admission prediction can even be achieved through simple Machine Learning algorithm such as a simple Linear Regression model, if for example we obtain the data such as GPA, GRE, LOR, SOP etc for previous admission applications in any particular university over the years and then craft a Linear Regression model using the past data. Now that we have that model we just need to pass

any student's GPA, GRE results, LOR and SOP through the model and the model should predict the chances of the student's admittance to the particular university [3]. But the precision or accuracy of this model will not be that high to be implemented in real life scenarios as it's based upon a very basic machine learning algorithm. Thus to increase the accuracy of the model various different machine learning techniques such as KNN, Random Forest, Support Vector Regression, Decision trees etc have been used by researchers.

Thus in this paper, we will be focusing on to these Machine Learning techniques in detail and try to understand their functioning and efficiency and review them to find out the best technique so far for the accurate prediction of admission in any university[1].

## II. LITERATURE REVIEW

There has been many papers on the topic of admission predictions but they all use varying methods using different machine learning approaches. Few of them uses some format of UCLA dataset which has the test scores which are more suitable for US universities admission for the admission prediction for training the data model, others use the academic merits, background and college admission criteria etc to achieve the prediction. Lets have a look at various research works.

[1] used the dataset form UCLA database and during their research the data had been downloaded 2000+ times and the dataset had parameters which were significant as they were used by the admission committees for granting admission. The dataset included GRE, TOEFL, graduate GPA, Statement of Purpose (SOP) and Letter of Recommendation (LOR). Then the scrapped data is cleaned and formatted to suit the needs of an Indian student. After the

segregation of the required data they have evaluated their dataset with 4 different machine learning algorithms namely 1. Linear Regression 2. Support Vector Regression 3. Random Forest and 4. Decision Trees and after the evaluation and calculations of errors and tabulation of the results they have chosen the best algorithm for their work.

There are other papers such as [3] which implement newer machine learning techniques such as Stacked Ensemble Learning.

The idea behind the application of this method dictates the training of several models and then combining their results to reach a solution for any given problem. Traditionally, the ensemble approach was used to merge weak algorithms together as when put together they prove efficient. Thus a stacked ensemble model is a model in which several machine learning algorithms are put together using meta-machine learning model, thus the predictions from multiple sub-models become the features of the meta-model increasing its efficiency. They collected data such as GRE scores, graduate GPA, TOEFL scores, number of publications, their program of application etc. In doing so a comparison with other supervised Machine Learning algorithms such as Linear Regression, Decision Trees, Random Forest, Naive Bayes Classifier, Logistic Regression etc according to their efficiency, has also been done in this paper and an observation has been made that this approach is more efficient than many previous university admission prediction models.

[6] have proposed a hybrid model using a combination of machine learning techniques to tackle the student admission predictions. The system proposed in the paper is built on top of a hybrid model of decision trees and neural networks. Thus the proposed system predicts the chances of admission in any specific university based on their academic merits, background and college admission

criteria. The prototype system then used live data of 2000+ high school student record and it suggests that the hybrid model of decision trees and neural networks performs better than any model which use only decision trees or neural network and also provides flexibility as the prediction provides the university based on the match of student's profile.

[2] Proposed a University Admission Predictor (UAP) system. They have tried to determine the predictions using various machine learning algorithms such as KNN, Decision Trees, Ridge Regression etc to predict the admission chances using the variables such as GRE, TOEFL, GPA, SOP, LOR etc. During the research they have tried to compare the algorithms mentioned above for choosing the best for their proposed UAP system. The proposed system is able to predict the admission chances with 79% of average accuracy.

[4] used data mining techniques to predict the performance of the students so that the predictions then can be used to support

Decision making for university admissions using Artificial Neural Networks (ANNs) and compared with other classification techniques such as Decision Trees, Naive Bayes and Support Vector Regression. The study tried to fill the research gap that could help universities finding the predictions about the academic performance of the students in the university prior to the admissions thus helping the universities shortlist the students who will perform better if given a chance at admissions and hence the university can gain better ratings as a result of better performing students. The system used the high school GPA, Scholastic Achievement Admission test score and General Aptitude Test score of 2039 students enrolled in computer science and information college of a Saudi public college for the dataset. The research also stated that the Scholastic Achievement Admission test score was the criteria which made

predictions with highest accuracy using Artificial Neural Networks which has an accuracy rate above 79% and thus ANNs outperforms the other classifiers such as Decision Trees, Naive Bayes and Support Vector Regression.

[5] tried to come up with a machine learning prediction system to predict the MS admission using the hypothetical UCLA data for the graduate admissions such as the GRE, TOEFL, GPA, SOP, LOR, Extra curricular activities, outstanding achievements and projects and research done by the applicant. The prediction system applied various machine learning algorithms such as Logistic Regression, Linear Discriminant Analysis (LDA), KNN, Decision Tree, Random Forest, Naive Bayes and Support Vector Machines (SVM) to find out the most accurate algorithm for the prediction. The author concludes that Logistic Regression (LR) and Linear Discriminant Analysis (LDA) performs best for this type of prediction with the dataset available and chose LR for the prediction model.

[7] proposed a Hybrid Recommender System for college admission prediction using data mining techniques and knowledge discovery rules. The HRSPCA system used a cascade of two recommender algorithms for the predictions. The first recommender is used to determine the track of a student for the preparatory exam and the second recommender assigns the specific college for the student if the student clears the preparatory exams. Thus in this way the HRSPCA system achieves better performance. The college predictor algorithm uses historical student GPA data along with academic merits, background, student records and admission criteria of the colleges to predict the colleges. The system was tested with live data of students from the database of the King Abdulaziz University (KAU).

[8] proposed a University Student Result Analysis and Prediction system by implementing the Decision Tree machine learning algorithm. The system can be used to predict the student results by analyzing their previous year/semester marks and the model also outlines the subjects in which the student is lacking. The system tries to help students find out the area or the subject where a student needs to put more focus so that he/she can avoid poor achievement and perform better. The data used for the prediction model is taken from the university database as well as through surveys done using google form. The data set contains information such as gender, extracurricular activities, number of tuition, programming skills, class test marks, assignment marks, attendance and previous GPA. The study tried to relate the performance prediction closely with CGPA. To train the data model weka tools was used. The objective of the study also revolves around the comparison of three different types of decision trees as well and concludes that the J48 decision tree algorithm outperforms the other two algorithms achieving the highest accuracy.

[10] built a web based college admission predictor system to reduce the time taken for the seat allotments after the entrance exams. The system tried to ease the problems of students as well as for the college administration. The system eases up the manual work needed for the seat allotments as well as helps in the computerization of the same. The web based system helps students decide their potential future colleges and after shortlisting they can apply to only those colleges where their chances of admission is more. The students need to register and feed their academic as well as their personal information such as marks, background details etc and then accordingly the system will predict their potential colleges. As the predictions come based on the entrance exams, the admin section is also maintained so that the prediction also reflects to the

admin and then after minimal manual verification and with increased ease due to computerization of the said subject the admin can allot the seat to the student. Thus the system enhances the speed and flexibility of admissions. The system uses generic algorithms for the computation and the website made for the system is made using PHP. The paper also touched on the basics of software development and tried to explain the implementation of the proposed system, such as White and black box testing with system and database design, modular design, input and output design etc.

[9] proposed a model for any engineering students academic performance prediction using various Linear Regression techniques as well as generic data mining techniques. The regression techniques discussed in the paper are Naive Bayes, Decision Tree, Support Vector Machine (SVM), Neural Networks, KNN etc, but the paper used Linear Regression, Logistic Regression and Multivariate Linear Regression for its prediction model. The paper proposed to build a machine learning model to predict the academic performance of students of engineering discipline. The model or the area of research revolves around the prediction of any student's performance in relation with his/her amount of time spent on the internet. That is the model tried to predict the outcomes based on the independent variables such as usage of internet for academic purposes, usage of internet for communication purposes, usage of internet for entertainment or social medial purposes, active duration on internet, usage of internet before semester exams etc, with the output being the cumulative grade point average or CGPA being the dependent variable throughout the research.

[11] presented an approach of designing and developing a recommender system for the students who wish to go for the graduate admissions in the

USA. The proposed system is used to help the student go for the graduate admissions in their preferred colleges based on their academic profile, hence it does a matchmaking process and determines the best college or university for a student. The proposed system used the means of database. It transformed the relational database for all the specific and relevant data of students into a universal database, I.e the database after transformation had the academic information of the students who got successful admissions in universities abroad. Then the proposed system with the help of an algorithmic model calculates similarities between the data of students who got successful admissions and the student who wants to get the admission based on the data provided by the willing student I.e the similarity is being check between the training data and the test data based on various weighted score using mean squared deviation metrics. The proposed system used KNN or K-nearest neighbour machine learning algorithm for the calculation of top K universities for the N similar users. The system uses various standardized test scores applicable in the USA such as GRE, TOEFL, IELTS as well as the academic profile of the student such as their CGPA in undergrad from the customized relation database to train and test the data set through the KNN algorithm to calculate the similarities. Thus after the prediction the system lists a number of universities the student can apply to for admissions with a better chances of getting accepted for admission.

[12] proposed a graduate school recommendation system for easing the troubles of students as the proposed system tries to help the students in decision making through shortlisting and recommending the appealing universities or colleges to the students according to their needs and requirements. The recommendation uses the Multi-Class Support Vector Machine (SVM) and K-Nearest Neighbour machine learning approaches. For the recommendation system,

the researchers have used SVM and KNN, SVM is chosen for its high accuracy with the complex linear and nonlinear data classification as the graduate data includes nonlinear type of data. With SVM, the core or the most appealing university is finalized based on the requirements of the student and the university/college through the fitness of classification. And after a core has been finalized, the KNN is used to predict a number of universities/colleges that are similar to the core defined by the SVM as the KNN is

very effective during computations of similarity based upon the nearness of the objects or classifiers which are other colleges/universities in this case. Thus through the collective use of SVM and KNN, this recommendation model promises better accuracy than other recommendation systems in the past. The data set used for this system is taken from online portals like Edulix.com [14] and Yocket.com [13] which archive data of various universities.

Table 1. Comparison of Algorithms based on Accuracy with respect to various research paper including precision ratio

S.No	Algorithms/ Methodologies used	Avg. Accuracy	Precision	F-measure	Reference Paper
1.	Linear Regression	72% [1], 79%[2]	72% [1]	72% [1]	[1], [2]
2.	Decision Tree	65% [3], 83% [8], 78% [4]	65% [3]	65% [3]	[3], [4], [8]
3.	Random Forest	66% [1], 77% [2], 62.5%[3]	63% [3]	63% [3]	[1], [2], [3]
4.	Support Vector Machines (SVM)	64% [1], 55% [3], 72% [4]	55% [3]	55% [3]	[1], [3], [4]
5.	Naive Bayes Classifier	53.3% [3], 73% [4]	54%[3]	47% [3]	[3], [4]
6.	Stacked Ensemble Learning	91% [3]	91% [3]	91% [3]	[3]
7.	Ridge Regression	78% [2]	NA [2]	NA [2]	[2]
8.	Artificial Neural Networks (ANN)	79.22% [4], 56% [3]	56% [3]	56% [3]	[4],[3]
9.	K-Nearest Neighbour	72\$ [2],	57% [3]	57% [3]	[3]



	(KNN)	57% [3]			
10.	Logistic Regression (LR)	55.88% [3]	54%[3]	54%[3]	[3]
11.	Linear Discriminant Analysis (LDA)	56% [3], 72% [5]	56% [3]	56% [3]	[3], [5]
12.	Hybrid Model Neural Networks + Decision Tree	80-90% [6]	80-90% [6]	80-90% [6]	[6]
13.	Hybrid Model SVM + KNN	55 % [12]	54% [12]	50% [12]	[12]

### III. METHODOLOGIES

Several machine learning algorithms have been used by several authors in the above mentioned literature review. In this section we will be seeing those methodologies and how these classifiers works. This section also deals with the inner workings of the listed methodologies.

#### Linear Regression

Linear regression is a regression model which uses supervised learning approach of machine learning. As the name suggests, a regression task is performed with this algorithm and a model is created based on the regression to derive meaningful contexts from a given set of data. It is one of the basic classifier used in machine learning. The algorithm is used to find out the relationship between dependent and independent variables. It is used to predict a dependent value or dependent variable (y) based on a given independent variable (x) thus finding a linear relationship between these two variables and thus named as Linear Regression[16,15].The linear regression model defines a variable which is based on a straight line described by equation  $Y = a + b x$  when one of the y-intersections of the line is y, and b is its pitch. First of all, the values of dependent variable Y and independent variable X are calculated with the aid of statistical methods by parameters A and B of the regression line. The regression line allows one to forecast the value of variable Y based on the value of

variable X independent. Thus, e.g. after a linear regression, the weight (dependent variable) of an individual from their height may be estimated (independent variable)[16].

#### Logistic Regression

Logistic regression is similar to linear regression but instead of a linear response given by the linear regression, the logistic regression has a binomial response variable. In the logistic regression we can have more than 2 continuous explanatory variables and it's easier to handle those variables simultaneously. In the case of more than one explanatory variable, logistic regression is used to calculate the odds ratio. With the fact that the outcome variable is binomial, the technique is very similar to multiple linear regression. The effect of each vector on the odds ratio of the observed occurrence of interest is the result. The biggest benefit is that when considering the relationship of all factors together, misleading consequences are avoided. A machine learning model created using logistic regression models the chance of all the outcomes based on the characteristics or features provided to the model through data feeding [17].

#### Decision Tree

The decision tree is the most effective and widely used classification and prediction method. A Decision tree is a tree structure that looks like a flowchart, with each internal node representing a test on an

attribute, each branch representing the test's result, and each leaf node (terminal node) holding a class name[15]. Decision trees are fundamentally a basic form of classifier, which is one of their advantages[18]. In its most basic form, we ask yes-or-no questions, and each internal node has a 'yes' and a 'no' child. Following the model from the topmost node, the root, to a node without children, a leaf, an item is sorted into a class depending on the responses that refer to the item in review[18]. Since they combine basic questions about the data in an intuitive manner, decision trees are often more easy to interpret than other classifiers such as neural networks and support vector machines. Methods for removing decision rules from decision trees have also shown to be effective. Unfortunately, minor differences in input data will also result in significant changes in the tree's structure. Decision trees are adaptable enough to deal with objects that have a combination of real-valued and categorical features, as well as items that lack any features[18]. Thus decision trees can produce rules that are easy to understand. Decision trees are used to build models without involving a lot of calculation and can accommodate factors that are both constant and categorical clearly showing which areas are most relevant for forecasting or classification[15].

### Support Vector Machines (SVM)

Support Vector Machines (SVMs) are a family of supervised learning methods that can be used to solve regression and classification tasks. It is most often used for sorting, but it can also be very useful for regression. A SVM classifier constructs a maximum-margin hyperplane in a transformed input space and divides the sample classes thereby minimizing the distance to the closest cleanly separated instances. This hyper-plane is nothing other than a line in two dimensions[15,19]. Each data object in the dataset is plotted in an N-dimensional space in SVM, where N represents the number of characteristics in the data.

Find the best hyperplane to isolate the results. As a result, SVM can only do classification task (i.e., choose between two classes). Even so, there are several methods for dealing with multi-class problems[15].

### Naive Bayes Classifier

The Naive Bayes classifier consist of a wide collection of algorithms which are based on Bayes' Theorem. Thus it's not a single algorithm but a family of algorithms in which all of them share a common principle, i.e every pair of a selected feature or characteristic is classified as independent of each other. That makes this algorithm a simple learning algorithm utilizing Bayes' rule with an assumption of all the attributes being conditionally independent of the class. Although in practice the independence assumption is often overlooked and violated, but still the classifier delivers competitive classification accuracy [15,20]. Nave Bayes is a technique for estimating the posterior likelihood  $P(y | x)$  of each class  $y$  given an object  $x$  using knowledge from sample data. We may use those projections for grouping or other decision support implementations until we have them[20]. The basic principle of Naive Bayes is that each function renders an independent and equivalent contribution to the result[15]. In mathematical terms, Bayes theorem find the probability of any given event's occurrence with respect to the probability of another event's occurrence.

### K-Nearest Neighbour (KNN)

K-Nearest Neighbors is a simple but important classification technique in Machine Learning. It is a supervised learning algorithm that is widely used in pattern recognition, data processing, and intrusion detection. It is easily applicable in real-world situations since it is non-parametric, which means it has no underlying assumptions regarding data delivery (as opposed to other algorithms such as



GMM, which assume a Gaussian distribution of the given data)[15,21]. The KNN algorithm presumes that identical objects happen nearby. In other terms, related objects are close together. The KNN algorithm is based on the premise that this assumption is valid for the algorithm to be useful. KNN captures the concept of similarity (also known as space, proximity, or closeness) with some mathematics we might have studied as children, such as measuring the distance between points on a line. Thus KNN works by calculating the distances between a query and all of the examples in the results, then choosing the number of examples (K) nearest to the query and voting for the most common symbol (in the case of classification) or averaging the labels (in the case of regression)[21]. It's simple to set up and grasp, but it has a big downside in being slightly slower as the scale of the data in use increases[15].

### Random Forest

A Random Forest is an ensemble learning method that can execute both regression and classification tasks by combining several decision trees and a technique known as Bootstrap and Aggregation, also known as bagging. The basic principle is to use several decision trees to determine the final production rather than relying on individual decision trees. Random Forest's foundation learning structures are multiple decision trees. We randomly select rows and features from the dataset to create sample datasets for each model. This section is known as Bootstrap[15]. So it is made up of a large number of independent decision trees that work together as an ensemble. Each individual tree in the random forest produces a class prediction, and the class with the most votes becomes the prediction of a model. The main point is the low correlation between models. Uncorrelated models can generate ensemble predictions that are more reliable than individual predictions. The explanation for this great result is that the trees shield each other from their individual

mistakes (as long as they don't all err in the same direction all of the time). While certain trees will be incorrect, several more would be accurate, allowing the trees to migrate in the same direction as a group. Thus a large number of relatively uncorrelated models (trees) acting as a committee will outperform all of the constituent models individually[22].

### Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are influenced by the way biological neural systems interpret information, such as the brain which is built of very complex webs of interconnected neurons. The information processing system is made up of several deeply integrated processing components (neurons) that collaborate to solve particular problems[23]. ANNs, just like people, learn by example[23]. The human brain is a tightly linked network of approximately  $10^{11}$ - $10^{12}$  neurons, with each connected neuron linked to an average of  $10^4$ - $10^5$  other neurons. So, on average, it takes about 10-1 for a human brain to make remarkably complicated choices. The aim of ANN models is to catch this kind of highly parallel computing based on distributed representations. ANNs are typically constructed from a densely interconnected series of simple units, each of which takes a number of real-valued inputs and generates a single real-valued output[15]. ANN learning, as learning in biological systems, requires changes to the synaptic interactions that occur between neurons[23]. Thus the electrical function of the brain and nervous system is simulated by ANN models. Processing components (also known as neurodes or perceptrons) communicate with one another. The neurodes are typically organized in a layer or vector, with the output of one layer acting as the entry to the next layer and potentially other layers. A neurode can be bound to all or a subset of the neurodes in the corresponding layer, simulating synaptic interactions in the brain. Weighted data signals

reaching a neurode mimic the electrical stimulation of a neuron and, as a result, information flow within the network or brain. The input data to a processing unit, in, are amplified by a connection weight,  $W(n,m)$ , which imitates neural pathway increasing in the brain. Learning is emulated in ANNs by adjusting the link strength or weights[23,24].

### Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a common pre-processing technique for dimensionality reduction problems in machine learning and pattern classification applications[25]. Since logistic regression is a classification algorithm, it has traditionally been restricted to two-class classification problems. Linear Discriminant Analysis is the favoured linear classification method where there are more than two groups. It is used to model group distinctions, such as dividing two or more classes. It is used to project functions from a higher dimension space to a lower dimension space[15]. We have two groups, for example, and they need to be effectively separated. Multiple features can be included in classes. It can result in any overlap using only one function to distinguish them, as seen in the figure below. Therefore, we will continue to increase the number of classification features[25,15]. To achieve maximum distance between means of two different classes and minimizing the variation within each class. The broad applications of LDA include Face Recognition, Medical classification of disease state as mild, severe, moderate based on diagnosis, Customer Type Identification etc.

### Stacked Ensemble Learning

Stacking is a means of grouping various classifications or models of regression. The most popular styles are Bagging or Boosting. There are several ways of ensembles. Bagging enables the average variation to decrease many related models

with high variances. The boost constructs some exponential versus small variance models to decrease the inclination[15,3]. Ensemble Learning is a form of machine learning technique, which training many weak models and combining the results with them to solve a specific problem[14]. An MLP alone is not the most effective model of the University Prediction Scheme neural network. There are a couple of more sophisticated profound learning processes. However, the results are extremely strong when you want to stack several MLP models to build a stacked ensemble model. To do so, we start by training the submodels of the MLP and then use the predictions of the submodels to train a metamodel. The outcome is a superior mode of deep learning[3]. Ensemble learning as a whole consists of the strategical combination of various learning models (such as supervised classification techniques) to create a superior model that can improve classification or estimation or even decrease opportunities to choose a bad outcome [3]. Ensemble learning model has 2 types. The first being Model Averaging Ensemble and second is Stacked Generalization. In model averaging, the projection better results, the forecasts of different models are merged. This may also be regarded as the disadvantage of the model, since there may be a case in which one variant is better than another but both weigh same. Hence, each model is allocated weights to address this situation. The best performing model is weighing than a poorer model by adding weights to each model. Finally, an average weighted and the last projection is estimated[3]. Whereas in stacked generalization or stacking the results of different classification algorithms are combined so that to make a clearer prediction with the help of meta classifier which is used to combine several models and at the end helps in generating the predictions through stacked ensemble learning model[3].

#### IV. CONCLUSION

The numbers of the students seeking further higher education abroad keeps increasing every year. And this number won't go down in upcoming years either. In our technological and competitive world the students constantly feel the need to upskill themselves and fare better than their peers and competitors everyday and thus after their under graduation, a vast number of students tend to leave for higher education abroad. To facilitate an easy migration, and to provide them with a tool which can help them shortlist their potential college/university based on various parameters will not just empower them in finding their preferred institutions but also ensure that it saves their precious time as well as their hard earned money at the same time. Many researchers have tried to come up with ideas using machine learning, data mining, and generic algorithm from time to time, each with their own benefits and drawbacks. If we go through the research done till date, the success rate of hybrid systems with the implementations of more than one algorithm concurrently is higher with low error rates, as the work done by each algorithm in any hybrid system relies on each of them only for the specific task and hence later collaboration ensures higher accuracy. In this regard, the stacked ensemble approach, which itself is based on an ensemble approach which tries to bring different algorithms together to gain better accuracy, fares better than all the other approaches discussed and reviewed in this paper. An improvement can also be made and the system can always be made more efficient and accurate using NLP or natural language processing methods which can help rate the quality of the written essays and paragraphs during the exams, so that those features can also help narrow down the chances of admissions over time.

#### V. REFERENCES

- [1]. M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.
- [2]. Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S, "Prediction for University Admission using Machine Learning" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-6 March 2020
- [3]. S. Sridhar, S. Mootha and S. Kolagati, "A University Admission Prediction System using Stacked Ensemble Learning," 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2020, pp. 162-167, doi: 10.1109/ACCTHPA49271.2020.9213205.
- [4]. H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," in IEEE Access, vol. 8, pp. 55462-55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
- [5]. Naman Doshi, "Predicting MS Admission", <https://medium.com/data-science-weekly-dsw/predictingms-admission-afbad9c5c599> February, 2018.
- [6]. S. Fong, Y. Si and R. P. Biuk-Aghai, "Applying a hybrid model of neural network and decision tree classifier for predicting university admission", 2009 7th International Conference on Information, Communications and Signal Processing (ICICS), Macau, pp. 1-5, 2009
- [7]. A. H. M. Ragab, A. F. S. Mashat and A. M. Khedra, "HRSPCA: Hybrid recommender system for predicting college admission," 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, India, 2012, pp. 107-113, doi: 10.1109/ISDA.2012.6416521.
- [8]. Md. Imdadul Hoque, Abul kalam Azad, Mohammad Abu Hurayra Tuhin, Zayed Us Salehin, "University Students Result Analysis and Prediction System by

- Decision Tree Algorithm" *Advances in Science, Technology and Engineering Systems Journal* Vol. 5, No. 3, 115-122 (2020) ,DOI: 10.25046/aj050315
- [9]. R.R.Rajalaxmi , P.Natesan , N.Krishnamoorthy ,S.Ponni,"Regression Model for Predicting Engineering Students Academic Performance" *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-7 Issue-6S3 April, 2019
- [10]. Annam Mallikharjuna Roa , Nagineni Dharani , A. Satya Raghava , J. Buvanambigai , K. Sathish,"College Admission Predictor" *Journal of Network Communications and Emerging Technologies (JNCET)* www.jncet.org Volume 8, Issue 4, April (2018)
- [11]. M. Hasan, S. Ahmed, D. M. Abdullah and M. S. Rahman, "Graduate school recommender system: Assisting admission seekers to apply for graduate studies in appropriate graduate schools," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2016, pp. 502-507, doi: 10.1109/ICIEV.2016.7760053.
- [12]. A. Baskota and Y. Ng, "A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018, pp. 277-284, doi: 10.1109/IRI.2018.00050.
- [13]. Yocket.com
- [14]. Edulix.com
- [15]. Geeksforgeeks.com
- [16]. Schneider, Astrid et al. "Linear regression analysis: part 14 of a series on evaluation of scientific publications." *Deutsches Arzteblatt international* vol. 107,44 (2010): 776-82. doi:10.3238/arztebl.2010.0776
- [17]. Sperandei, Sandro. "Understanding logistic regression analysis." *Biochemia medica* vol. 24,1 12-8. 15 Feb. 2014, doi:10.11613/BM.2014.003
- [18]. Kingsford, Carl, and Steven L Salzberg. "What are decision trees?." *Nature biotechnology* vol. 26,9 (2008): 1011-3. doi:10.1038/nbt0908-1011
- [19]. Shmilovici A. (2005) Support Vector Machines. In: Maimon O., Rokach L. (eds) *Data Mining and Knowledge Discovery Handbook*. Springer, Boston, MA. [https://doi.org/10.1007/0-387-25465-X\\_12](https://doi.org/10.1007/0-387-25465-X_12)
- [20]. Webb G.I. (2011) Naïve Bayes. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30164-8\\_576](https://doi.org/10.1007/978-0-387-30164-8_576)
- [21]. Zhang, Zhongheng. "Introduction to machine learning: k-nearest neighbors." *Annals of translational medicine* vol. 4,11 (2016): 218. doi:10.21037/atm.2016.03.37
- [22]. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- [23]. Grosan C., Abraham A. (2011) Artificial Neural Networks. In: *Intelligent Systems. Intelligent Systems Reference Library*, vol 17. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-21004-4\\_12](https://doi.org/10.1007/978-3-642-21004-4_12)
- [24]. <https://www.sciencedirect.com/topics/engineering/artificial-neural-network>
- [25]. Tharwat, Alaa & Gaber, Tarek & Ibrahim, Abdelhameed & Hassanien, Aboul Ella. (2017). Linear discriminant analysis: A detailed tutorial. *Ai Communications*. 30. 169-190,. 10.3233/AIC-170729.

#### Cite this article as :

Prince Golden, Kasturi Mojesh, Lakshmi Madhavi Devarapalli, Pabbidi Naga Suba Reddy, Srigiri Rajesh, Ankita Chawla, "A Comparative Study on University Admission Predictions Using Machine Learning Techniques", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 7 Issue 2, pp. 537-548, March-April 2021. Available at  
doi : <https://doi.org/10.32628/CSEIT2172107>  
Journal URL : <https://ijsrcseit.com/CSEIT2172107>