DB10

by Vamshikrishna Namani

Submission date: 09-Mar-2023 04:13PM (UTC+1000)
Submission ID: 2032792831
File name: conference-template-a4_1_2_1.docx (458.89K)
Word count: 2320
Character count: 13448

STUDENT ADMISSION PREDICTION

Abstract

The student admissions prediction problem involves using data from past applicants to develop a model that can accurately predict whether a new applicant will be accepted or rejected by an educational institution. In this study, we will examine several factors that may influence admissions decisions, such as academic performance, standardized test scores, extracurricular activities, and demographic characteristics. We will use machine learning techniques to develop a predictive model and evaluate its performance on a set of test data. This research has important implications for educational institutions looking to optimize their admissions process, as well as prospective students looking to improve their chances of gaining admission.

Introduction

The Student Admission Prediction Project is a data-driven initiative to develop a predictive model to predict a student's chances of admission based on their academic profile and other relevant factors. The Student Admissions Forecasting Project uses data -driven insights to increase transparency and fairness in the college admissions process and make it easier to match qualified students with the right universities and colleges. Student admission prediction is a data-driven approach to predicting a student's likelihood of being admitted to a particular college or university based on their academic profile and other relevant factors. This approach involves collecting and analyzing large amounts of data on previous applicants and their admissions outcomes, as well as data on current students and their academic performance.

Literature Review

- Predicting Factors: Various studies have identified several predictors that help students make admissions decisions. For example, academic records such as GPA and standardized test scores are often used as primary predictors. Other factors such as extracurricular activities, personal statements, and letters of recommendation were also found to be significant predictors of admission.
- Performance measures: The study used various performance measures to assess the accuracy and performance of the predictive models.
- Data preprocessing: Data preprocessing is a critical step in developing accurate predictive models. The research used various techniques such as feature selection, normalization and coding to prepare the data for analysis.
- Model interpretability: Model interpretability is an important factor in ensuring that predictions are transparent and understandable.

The research used a variety of techniques, including attribute, importance analysis and decision tree visualization, to explain patterns, and understand factors that influence admissions decisions.

By using machine learning algorithms and other statistical techniques to analyze this data, admission prediction models can be developed that are able to identify patterns and trends that can help predict which students are most likely to be admitted. This can be especially helpful for students who are applying to highly competitive schools, as it can help them make more informed decisions about where to apply and increase their chances of being admitted.

In general, literature on student admissions predictions shows that machine learning algorithms can effectively predict admissions outcomes. However, the accuracy and performance of the model may vary depending on the data and the problem to be solved. Additionally, interpretability and transparency are key factors to ensure reliable and trustworthy forecasts.

EXISTING SYSTEM

Students can predict their next admission chance by doing mathematical calculations on previous scores. It takes a long time and is difficult to plan by hand. Predicting admission chances based on previous scores using mathematical calculations can be a useful tool for students who are applying to colleges or universities. However, doing these calculations manually can be time-consuming and difficult to plan, especially for students who are applying to multiple schools or who are considering different academic programs or majors.

proposed system

The proposed student admissions prediction system would use machine learning algorithms to analyze historical admissions data and identify patterns and correlations between various factors and admissions outcomes. The system will consider factors such as academic achievement, test scores, and other relevant factors that may affect a student's chances of admission. The system will use a multiple random forest regression algorithm. The algorithm provides output by processing different properties.

Dataset and Visualization

- The dataset includes features like GRE scores, TOEFL scores, undergraduate GPA, research experience, and admissions results.Once an appropriate data set is available, data visualization techniques can be used to explore the data and identify patterns and correlations between different characteristics and admission outcomes. This can be done using tools like scatterplots, histograms, and heatmaps. Data visualization can also be used to identify outliers or missing data that may need to be addressed before developing a predictive model.
- Data visualization is the process of representing complex data and information in a visual format that is easier to understand and interpret. This can include creating charts, graphs, maps, and other visual representations that help to communicate key trends and patterns in the data.

Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit	
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...	...
395	396	324	110	3	3.5	3.5	9.04	1	0.82
396	397	325	107	3	3.0	3.5	9.11	1	0.84
397	398	330	116	4	5.0	4.5	9.45	1	0.91
398	399	312	103	3	3.5	4.0	8.78	0	0.67
399	400	333	117	4	3.2	4.0		1	

Fig. 1. Dataset

Preprocessing

- Pre-processing is an important step in developing an accurate and efficient student admission prediction model. Some key steps in pre-processing include:

1) Data cleaning: This involves identifying and addressing missing data, addressing outliers, and resolving any inconsistencies or errors in the data.

2) FeatureSelection : It is important to select the most relevant characteristics that are likely to have an impact on admission results. This may involve the use of techniques such as correlation analysis or feature importance ranking.

3) Data Normalization: This involves scaling data so that it has a consistent range and distribution.This is important because different features can have different scales, and normalization can help ensure that all features.

4) Data encoding: One of the most common techniques for categorical encoding is one-hot encoding or dummy variable encoding. In this technique, each category in a categorical variable is represented as a separate binary variable, with a value of 1 if the observation falls into that category, and 0 otherwise. For example, if we have a categorical variable "Color" with categories "Red," "Blue," and "Green," we would create three binary variables, "Color_Red," "Color_Blue," and "Color_Green," and assign them values of 1 or 0 depending on the color of each observation. Overall, categorical encoding is an important step in preparing data for machine learning algorithms. By transforming categorical variables into a numeric format, we can use them as inputs to algorithms such as linear regression, decision trees, and neural networks, allowing us to build accurate and effective predictive models.

5) Data splitting: Data splitting is a technique used in machine learning to divide a dataset into two or more subsets for the purpose of training and evaluating predictive models. The most common type of data splitting involves dividing the dataset into two subsets: a training set and a test set.

The training set is used to develop the predictive model. This involves training the model using various machine learning algorithms such as linear regression, decision trees, and neural networks. The goal of the training phase is to develop a model that accurately predicts the outcome variable based on the input variables.

In general, preprocessing is a critical step in developing an accurate and efficient student enrollment prediction model. By cleaning, selecting, normalizing, encoding and partitioning the data, we can ensure that the model is based on high quality data and can accurately predict the admission outcome of freshmen.

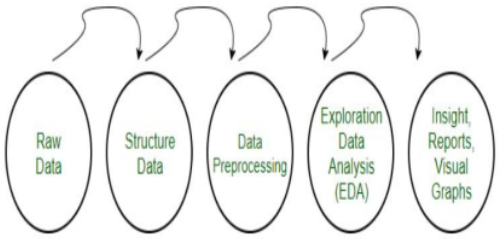


Fig. 2. Data Pre-processing

System Analysis

This analysis includes identifying factors that influence admissions decisions, determining what data needs to be collected to make accurate predictions, and designing algorithms or models that can use that data to make predictions.
Here are the steps typically involved in an analysis of a student admissions forecasting system:

- ✓ Identifying Factors Affecting Admissions Decisions: The first step is to identify the factors that typically affect admissions decisions .
- ✓ Data Collection: Once the factors influencing admission decisions have been identified, the next step is to collect data on these factors. This data may be collected from a variety of sources, including student applications, test scores, transcripts, and other documents.
- ✓ Data cleaning and pre-processing: After data is collected, it may need to be cleaned and pre-processed to remove any errors or inconsistencies. This may involve identifying missing data points, correcting errors, and normalizing data.
- ✓ Designing Algorithms or Models: After preprocessing the data, the next step is to design algorithms or models that can use that data to make predictions. This may involve choosing an appropriate machine learning or statistical technique to analyze the data and then training a model using a subset of the data.
- ✓ Model Deployment: Once a model has been tested and validated, it can be deployed to predict student admission decisions.

In general, systematically analyzing student admissions predictions requires an in-depth understanding of the factors that influence admissions decisions, as well as the statistical and machine learning techniques needed to analyze the data and make accurate predictions . By following these steps, schools and colleges can improve their admissions process and ensure that they admit the most qualified and deserving students.

Statistical and machine learning techniques are needed to analyze the data and make accurate predictions. These techniques include data cleaning, feature engineering, model selection, and model evaluation. Data cleaning involves removing errors and inconsistencies from the data, while feature engineering involves transforming the data into a format suitable for analysis. Model selection involves choosing the best predictive model for the data, while model evaluation involves assessing the accuracy of the model.

By following these steps, schools and colleges can improve their admissions process. For example, they can use predictive models to identify the most qualified and deserving students, and allocate resources more effectively. They can also use the models to identify the factors that are most important in the admissions process, and use this information to improve their admissions criteria.

In this project, I build a linear regression model to predict the chance of admission into a particular university based on studetent's profile.

Instructions for Input Features

- GRE Score (out of 340)
- TOEFL Score (out of 120)
- University Rating (out of 5)
- Statment of Purpose {SOP} (out of 5)
- Letter of Recommendation {LOP} Strength (out of 5)
- Undergraduate CGPA (out of 10)
- Research Experience (Either 0 or 1)

Fig. 3. Dataset Instructions

Methodology and Implementation

Ensemble methods are machine learning techniques that combine multiple weak classifiers to create a strong classifier. The goal of ensemble methods is to improve the accuracy of the model by reducing the errors of individual classifiers. In this case, the authors used four different ensemble methods: voting classifier, bagging, AdaBoost, and stacking.

The authors implemented a speech classifier on a dataset using previously implemented classifiers: decision trees, support vector machines (SVC), and logistic regression. The voice hyperparameter chosen was Hard Voice, which is used when SVC does not provide a probability measure.

The accuracy of the voice classifier was found to be 0.91, which is a good result. Logistic regression and SVC are powerful classifiers that use linear models to classify data. However, when comparing the accuracy of the speech classifier with the other ensemble methods, the accuracy increased.

Bagging is an ensemble method that uses a classifier to randomly sample instances with replacement. The authors used the Bagging method from the MLXTEND library with four SVC classifiers, Random Forest, Gaussian Naive Bayes, and Metaclassifier Logistic Regression.

The stacking classifier is an advanced ensemble method that uses multiple models as base classifiers, and then trains a meta-classifier on the outputs of the base classifiers. After training and testing the stacking classifier on the test data, the accuracy of the classifier reached a value of 0.95, which is

considered better in terms of accuracy compared to the other ensemble methods and all other classifiers.

In summary, the authors used ensemble methods to improve the accuracy of the speech classifier. The stacking classifier was found to be the most accurate among all classifiers and ensemble methods. This highlights the effectiveness of using ensemble methods in machine learning to improve the accuracy of predictive models.

Result and Analysis

When analyzing the results of a student admissions prediction system, it is important to use metrics such as accuracy and recall to assess the performance of the model used. Accuracy measures the proportion of correct predictions made by the model, while recall measures the proportion of true positives that the model correctly identified. By using these metrics, schools and colleges can determine how well the model is performing and make any necessary adjustments to improve its accuracy and efficiency.

Additionally, it is important to examine the factors identified by the model as being most important in predicting admission decisions. By understanding these key factors, schools and colleges can make better decisions about which students to admit and improve the admissions process overall. This information can also be used to identify any biases in the admissions process and make adjustments to ensure that all qualified students have an equal opportunity to be admitted.

Overall, analyzing the results of a student admissions prediction system involves a combination of statistical and machine learning techniques, as well as a thorough examination of the factors that influence admissions decisions. By using this approach, schools and colleges can improve their admissions process, admit the most qualified and deserving students, and ensure fairness and transparency in the admissions process.

For example, if a model is used to predict whether a student will be admitted to a university, precision measures the proportion of students who were admitted out of all the students that the model predicted would be admitted. Precision is a useful metric when avoiding false positives is more important than identifying all positive cases.

When analyzing the results of a predictive model for student admissions, it is important to use these metrics to evaluate the performance of the model. In addition, it is important to examine the factors that the model identified as most important in predicting admissions decisions. This information can be used to improve the admissions process and make more informed decisions about which students to admit.

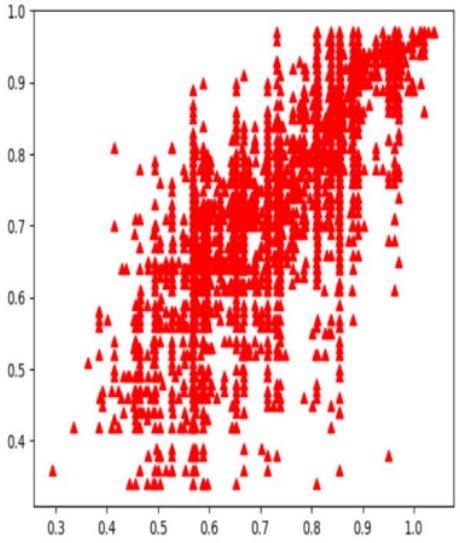
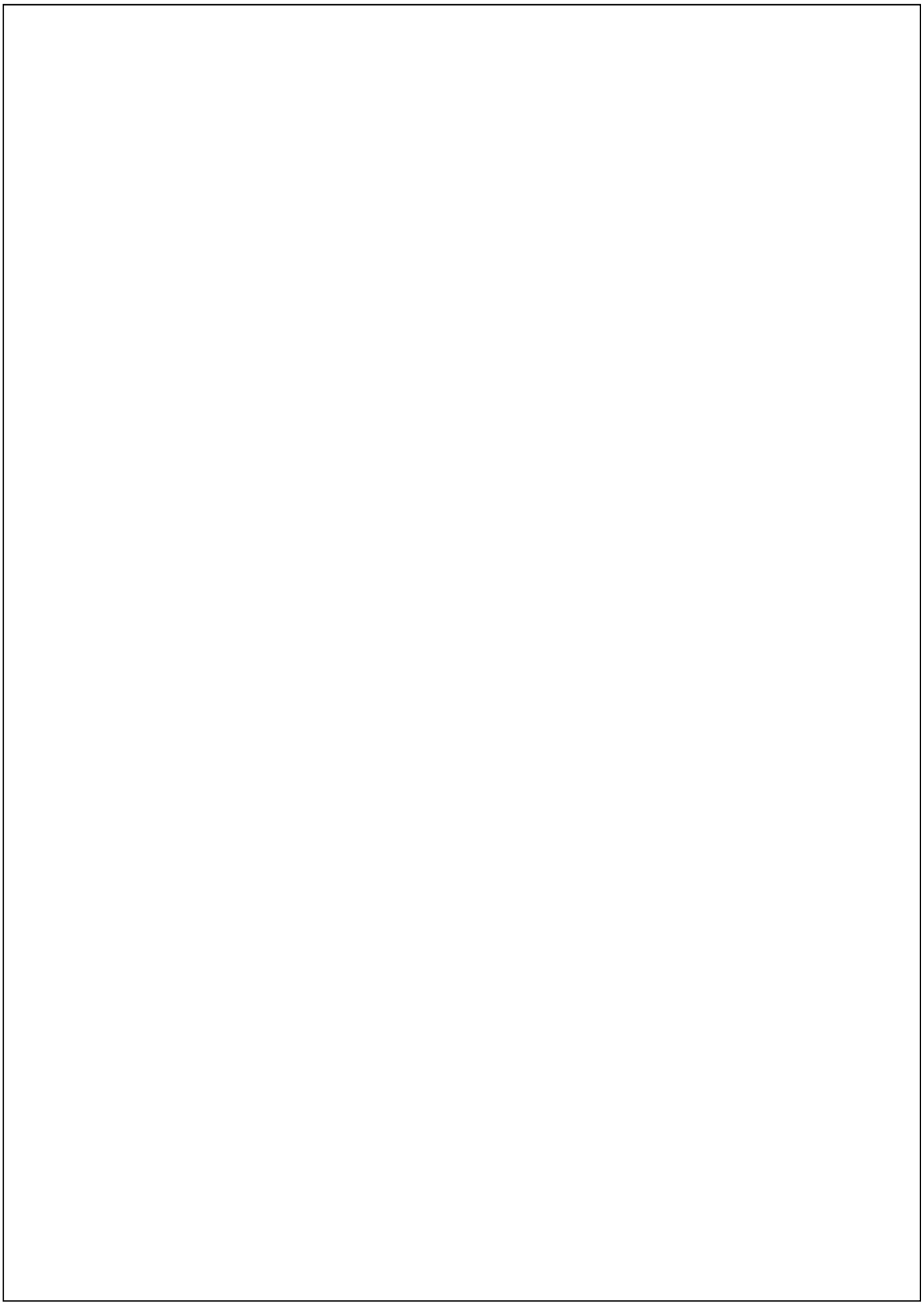


Fig. 4. Predicted values and tested values

Conclusion

The student admissions prediction problem involves using data from past applicants to develop a model that can accurately predict whether a new applicant will be accepted or rejected by an educational institution. In this study, we will examine several factors that may influence admissions decisions, such as academic performance, standardized test scores, extracurricular activities, and demographic characteristics.

This research has important implications for educational institutions looking to optimize their admissions process, as well as prospective students looking to improve their chances of gaining admission.



DB10

ORIGINALITY REPORT

6%	3%	3%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Michael Onyema Edeh, Osamah Ibrahim Khalaf, Carlos Andrés Tavera, Sofiane Tayeb et al. "A Classification Algorithm-Based Hybrid Diabetes Prediction Model", Frontiers in Public Health, 2022 Publication	1%
2	business-analytics-courses-in-hyd.blogspot.com Internet Source	1%
3	Ab Shaqoor Nengroo, K.S. Kuppusamy. "Machine learning based heterogeneous web advertisements detection using a diverse feature set", Future Generation Computer Systems, 2018 Publication	1%
4	www.ncbi.nlm.nih.gov Internet Source	1%
5	Submitted to University of Essex Student Paper	1%
6	Submitted to University of Northampton Student Paper	1%

7	Submitted to Taylor’s Education Group Student Paper	1 %
8	Submitted to University of Surrey Student Paper	1 %
9	Submitted to Berlin School of Business and Innovation Student Paper	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		