Classification of Admission Data Using Classification Learner Toolbox

Dr.M.Senthamil Selvi¹, S.Jansi Rani²

¹ Prof/IT, Sri Ramakrishna Engineering College, Coimbatore
² AP/IT, Sri Ramakrishna Engineering College, Coimbatore
Email: ¹senthamilselvi@srec.ac.in; ²jansi.sankar@srec.ac.in, jansiranis89@gmail.com

Abstract. The analysis of data useful for further process and find the usefulness for the futher. In education sector, Admission is hottest topic. For the counselling of the admission in reputed institution or foreign university, the past data of the person are taken into the consideration. In this Paper, Admission data taken from the reputed Website (Dataset Weblink) and conduct the study of classification methods like Naïve bayes and its accuracy using Matlab. The performance of different classification methods from MATLAB classifier learner toolbox for admission data. Finally, the paper choosen the suitable classification method for the admission data set.

Keywords- Classification Methods, Naïve bayes, Admission data, Matlab

INTRODUCTION
Every year, many students applying for the admissions in their own country and foreign universities especially for Post graduate degree and research studies. There are many admission related websites and statistics are available during admission season and no website does not predict the individual admission data. With the help of machine learning, many real world applications are booming and predicting with high accuracy. The same can be implemented to train the admission dataset and predict the same with highest accuracy.

The aim of the paper has three folds:
(1) Information about the dataset
(2) Exploratory Data Analysis (EDA)
(3) Load the data set using MATLAB and compare the model using Classifier App and choose the best suitable model for admission data set by considering the methods accuracy
(4) Test the Model (high accuracy model)

The paper has organized as follows: Section2 describes the literature survey on Prediction of student get an admission or not. Section5 describes the MATLAB process, Section4 tells which model suitable for admission dataset based on the accuracy.

LITERATURE REVIEW
For predict the admission need lot of energy and experience in Education Filed. The paper [1] studied the graduate admission process in American universities and used a machine learning approach. The goal of the authors were build a decision support model that allows candidates to make decisions on which universities to apply to, chances of admission, and a slew of other decision-related issues.
presented a system that can achieve high accuracy. The author proves that a mixture of approaches can provide better results than any of the individual approaches.

The Paper [4] presented a two-stage statistical model that first predicts the acceptance of an applicant to a selective liberal arts college and later predicts the enrollment of that accepted student.

This paper implement the different classification algorithm in MATLAB with the help of Classifier Learners Toolbox, choose the best and predict the admission dataset.

INFORMATION ABOUT THE DATASET:

Dataset used for the analysis is Admission dataset which consists of 9 columns and 401 rows consists of GRE score, TOEFL score, Statement of Purpose, Letter of Recommendation, Academic performance – GPA, Research and Chance of Admission.

The research and chances of admission are binary value like 0 or 1 values.

The outcome of the data is that whether the person has a chance to get admission. So, the problem is binary classification problem.

![Admission Dataset](image)

Fig 1: Admission Dataset

Fig 1 gives details about Admission Dataset (Table shown only first twenty rows)

EXPLORATORY DATA ANALYSIS (EDA):

EDA explore the data and understand the features of the given dataset. This dataset has a binary response variable called as chance of admit (i.e., Outcome response variable). The variable rank of the university takes the values from 1 to 4. Institute with high ranked as 1 and 4 for low ranked institutions.

IMPLEMENTATION USING MATLAB

To implement algorithm for the admission dataset, the MATLAB 2019b and Classification Learner toolbox was used. We need to classify into 2 different classes according to the given scenarios.
1. Dataset loaded into MATLAB environment with the following command.

```
>> load('admissionData.mat')
```

![Fig 2: Dataset in MATLAB](image)

Fig 2 shows the loaded Admission dataset into MATLAB

2. Open a Classifier App under Machine Learning and Deep Learning tab.

![Fig 3: Classifier Session](image)

3. In classifier App [3], start the new session and load the admissiondata and select the required variable classification models. Chance of admit is response variable for the current admissiondata. In new session, Workspace variable choosed as admissiondata and response as chance of admission as 0 or 1. Also, remove any no of response variable for implementation. The selected validation is hold-out method and the folds are increased as 10 (default hold is 5). Finally click start session tab.

4. Classification:
The classifier Naïve Bayes are collection of classification algorithms based on bayes theorem (i.e, dependent and independent variable)
Model 1: In Gaussian Naïve Bayes, a continuous values associated with each feature are assumed to be distributed according to Gaussian distribution.

![Model 1](image)

**Fig 4:** Accuracy and Confusion Matrix for Model 1

Fig 4 shows that accuracy of the Model 1 is 86.0% and confusion matrix is displayed. The confusion matrix used to analyze the TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) values. In the Model 1, TP is 315 values are labelled 1 as 1. TP rates for 1 and 0 are 86% and 83% respectively and FN are 14% for 1 and 17% for 0.

Model 2:
Kernel Naïve Bayes [2] implemented for numerical attribute. The non-parametric estimation procedures utilize the weight function is known as kernel. For the estimation of conditional expectation of a random variable in the dataset, Kernel regression mechanism is used.

![Model 2](image)

**Fig 5:** Accuracy and Confusion Matrix for Model 2

Fig 5 shows, accuracy for Kernel Naïve Bayes is 90% and TP rate is increased as 92% for 1 as 1.

Model 3: Gaussian Model with PCA on
PCA (Principal Component Analysis) used to reduce the dimension of the dataset. This is one of the data reduction technique.
Fig 6: Accuracy and Confusion Matrix for Model 3

Model 4: Linear SVM with PCA on

Fig 7: Accuracy and Confusion Matrix for Model 4

Model 5: Linear SVM
6. Comparison for Model 1 to Model 5

| Model                                      | Accuracy | Total Misclassification Rate | Prediction Speed | Training Time |
|--------------------------------------------|----------|------------------------------|------------------|---------------|
| Model 1- Gaussian Naïve Bayes              | 86%      | 56                           | ~3700 obs/sec    | 50.855 sec    |
| Model 2- Kernel Naïve Bayes                | 90%      | 40                           | ~2200 obs/sec    | 4.7022 sec    |
| Model 3- Gaussian Naïve Bayes with PCA on  | 91.3%    | 35                           | ~2100 obs/sec    | 5.2655 sec    |
| Model 4- Linear SVM with PCA on            | 91.3%    | 35                           | ~2500 obs/sec    | 6.6781 sec    |
| Model 5- Linear SVM                        | 93.3%    | 27                           | ~8200 obs/sec    | 1.3233 sec    |

**Table 1:** Model Result

Model results table: 1 shows the different parameters of the models such as Accuracy, Total Misclassification rate, Prediction Speed and Training Time. Based on all parameters of the Model, Model 5- Linear SVM model gave better accuracy (93.3%) for the admission dataset. The ROC curve for Linear SVM, the curve based on TP rate against FP rate.
7. Predict Using Model:

The Linear SVM is best model based on accuracy and the same is exported using Export Model Tab and predict the data.

![Export Model](image)

**Fig 10:** Export Model

The chances of admit are NaN and finally predicted as 1.

```
>> yfit=trainedModel.predictFcn(data)
```

**Fig 11:** Predict Function of Trained Model

8. Conclusion:
The main result of the admission dataset was done with the help of MATLAB classifier learner toolbox. Different Classification algorithms were tested for admission dataset. Based on the result given in the table 1, Linear SVM model produced better result in terms of accuracy (93.3%) and Misclassification rate. Finally, Linear SVM model exported and tested the admission dataset with prediction as 1 or 0.

References:
[1] Narender Gupta, Aman Sawhney, Dan Roth; “Will I Get in? - Modeling the Graduate Admission Process for American Universities” in proceedings of 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)
[2] Dishant Khanna, Arunima Sharma, “Kernel – Based Naïve Bayes Classifier for Medical Predictions” in Advances in Intelligent Systems and Computing book series (AISC, Volume 695)
[3] Tomáš Pokorny ; Jan Tesarik, “Microwave Stroke Detection and Classification Using Different Methods from MATLAB’s Classification Learner Toolbox”, in proceeding of 2019 European Microwave Conference in Central Europe (EuMCE).
[4] Prof. Romi Morzelona. (2019). Histogram Based Data Cryptographic Technique with High Level Security. International Journal of New Practices in Management and Engineering, 8(04), 08 - 14.
[5] Bruggink, T. H., & Gambhir, V. (1996). Statistical models for college admission and enrollment: A case study for a selective liberal arts college. Research in Higher Education, 37(2), 221–240. doi:10.1007/bf01730116.
[6] Natarajan, B., Obaidat, M.S., Sadoun, B., Manoharan, R., Ramachandran, S. and Velusamy, N., 2020. New Clustering-Based Semantic Service Selection and User Preferential Model. IEEE Systems Journal. DOI: 10.1109/JSYST.2020.3025407.
[7] Nataraj, S.K., Al-Turjman, F., Adom, A.H., Sitharthan, R., Rajesh, M. and Kumar, R., 2020. Intelligent Robotic Chair with Thought Control and Communication Aid Using Higher Order Spectra Band Features. IEEE Sensors Journal, DOI: 10.1109/JSEN.2020.3020971.
[8] Babu, R.G., Obaidat, M.S., Amudha, V., Manoharan, R. and Sitharthan, R., 2020. Comparative analysis of distributive linear and non-linear optimised spectrum sensing clustering techniques in cognitive radio network systems. IET Networks, DOI: 10.1049/iet-net.2020.0122.
[9] Sitharthan, R., Yuvaraj, S., Padmanabhan, S., Holm-Nielsen, J.B., Sujith, M., Rajesh, M., Prabaharan, N. and Vengatesan, K., 2021. Piezoelectric energy harvester converting wind aerodynamic energy into electrical energy for microelectronic application. IET Renewable Power Generation, DOI: 10.1049/rpg2.12119.
[10] Sitharthan, R., Sujatha Krishnamoorthy, Padmanaban Sanjeevikumar, Jens Bo Holm-Nielsen, R. Raja Singh, and M. Rajesh. "Torque ripple minimization of PMSM using an adaptive Elman neural network-controlled feedback linearization-based direct torque control strategy." International Transactions on Electrical Energy Systems 31, no. 1 (2021): e12685. DOI: 10.1002/2050-7038.12685.