KASU JOURNAL OF MATHEMATICAL SCIENCES (KJMS) VOL. 2, ISSUE 2, December 2021
ISSN 2734-3839 (Print), ISSN 2735-962X (E-Copy)
http://www.journal.kasu.edu.ng/index.php/kjms

KASU JOURNAL OF MATHEMATICAL SCIENCES

Analysis of Students' Academic Performance Using Traditional And Machine Learning Classifiers

Olalude Gbenga A[1], Amusan Ajitoni S[2], Adeshina Ibrahim O[3]

[1,2,3]Department of Statistics, Federal Polytechnic Ede, Osun State, Nigeria
Email; Olalude.gbenga@federalpolyede.edu.ng1; Adeshinaibrahim10@gmail.com[3]

Abstract

Unlike traditional statistical techniques of analyzing data, machine learning algorithms have made it easier to feed computer with (partitioned) dataset then slightly program it in order to obtain a model with the most precise classification. This study juxtaposed the classification and predictive performance of traditional statistical classifiers as compared with machine learning algorithms. Simultaneously, the study aim to classify and predict students' academic performance in Federal polytechnic Ede, Osun state. The considered classifiers include Logistic regression, Decision Tree, support vector machine and Naive Bayes. Students' academic performance was classified as a factor of other variables which include; Age at admission, Years before admission, Program type, High school type, 0'level, and Gender. These classifications were subjected to varying train-test ratio ranging from 90-10 to 50-50. It was obtained that some of our classifiers performs below 50% accuracy, however decision tree algorithm yielded the most precise classification has it outperforms both Naïve Bayes and logistic regression largely, but support vector machine slightly. It is highly evident that all classifiers yielded maximum individual accuracies in 90-10 train-test ratio. Thus, we conclude that the decision tree algorithm under 90% training set and 10% testing set is appropriate for future prediction of students' academic performance and it is recommended that it should be adopted for similar studies.

Keywords: Classification, machine learning, support vector machine, decision tree, Naïve Bayes, Logistic regression

1. Introduction

Predicting students' academic performance is an integral part of an education system, as the overall growth of the education system is directly proportional to the success rate of the students in their examinations. Therefore there are many situations where the performance of the students' are necessary to be predicted, for example to identify eligible students for participating in placement activities, to identify students eligible for scholarships and to find the weak students so that remedial action can be taken for their betterment. Often times, students' characteristics or experience in a course might determine if they will pass or fail (Kweon, Ellis, Lee and Jacobs, 2017). However this set of characteristics are considered to be personal which work in line with other external (e.g social) factors. Justifying from the point of view of these characteristics, one might model these characteristics to serve as basis of determining the possible outcome of a student after his study activities using statistical classification techniques. With the advancement of technology in recent days, machine learning which as a branch of data mining is becoming more popular than the traditional techniques of data analysis (e.g. regression analysis, monte carlo) for

its effectiveness in decision making processes and model formulation. Traditional data technique is a manual statistical procedure being programmed to automate a statistical technique. This indicates that without anyone programming the logic, one has to manually formulate or code rules. On the other hand, machine learning is an automated process which increases the value of an embedded analytics in many areas, including data prep, natural language interfaces, automatic outlier detection, recommendations, causality and significance detection. All of these features help speed users insight and reduce decision bias.

Not to exaggerate, authors like (Xu, Wang, Peng and Wu, 2019) and (Sokkhey & Okazaki, 2020) have lauded the predictive performance of machine learning algorithms as compared to the traditional statistical techniques. They established the fact that machine learning gives room for the computer to learn from the available data (training data) until it return an optimum model for predictive actions on the unseen data (test data). Without any form of prejudice, traditional data analysis techniques are quite easy to approach for a lay-man but not really a good fit when it comes to big data analysis. Despite this advantage, traditional data analysis techniques might fail to render the most precise result based on other data characteristics e.g. imbalance dataset.

Considering the complexity in understanding the concept of machine learning and the ease with the use of traditional data analysis technique, one might be tricked to opt for the latter in expense of the former. However the likes of (Dekker, Tomic, Tomic, Rosenberg-Hasson, Maecker and Davis, 2019), (Wolff, O'Donncha and Chen, 2020) and (Asif, Martiniano, Vicente and Couto, 2018) has made it clear the margin of advantage that machine learning algorithms have over tradition data analysis techniques. Similarly, Churpek, Yuen, Winslow, Meltzer, Kattan and Edelson (2016) who compared the performance of several machine learning algorithms (Random forest, Gradient boosted machine, Bagged trees, support vector machine, Neural network, Decision tree and K nearest neighbor algorithms) relative to logistic regression, using observational cohort dataset from hospitalized ward patients for detecting clinical deterioration on the wards in a large, multicenter database has found that several machine learning methods perfectly predicted clinical deterioration compared to the logistic regression model. In relation to our study, Cortez and Silva (2008) have previously reviewed the prediction of secondary school students' performance in two essential courses (mathematics and Portuguese) using their previous score in the prior sessions and other demographic factors and employed four data mining methods of Decision trees, Random Forests, Neural networks and Support Vector machines approach. They obtained a result which indicate that the prediction was attainable provided the grades of the previous session were known. This endorses that the prediction of students' performance is premised on previous performance and hence indicates that a student's performance is closely related to the performance in previous course (most likely a prerequisite course). Another similar study by Sembiring, Zarlis, Hartama, Ramliana & Wani (2011), showed that machine learning capabilities provided effective improving tools for student performance. The study further showed how useful machine learning can be in higher education particularly to predict the final performance of student. The researchers collected data from students by using questionnaire to find the relationships between behavioral attitude of student and their academic performance after which they applied machine learning algorithm

decision tree and Support Vector Machine (SVM) to predict the students' final grade. Also the students were clustered into groups using kernel k-means clustering. They pinpointed from their model that there is a strong correlation between mental condition of student and their final academic performance. To combine the theme of all previous efforts, three machine learning algorithm were applied in addition to logistic regression classifier which will serve as the relative group with which comparison will be made with our other models and the most precise algorithm will serve as basis of our future students' performance classifier.

2. Methodology

2.1 Data Source and Method of Collecting Data

Record of students that was obtained include: social, demographic and other educational records of students from the department of statistics, federal polytechnic Ede, Osun state. The obtained data comprise Matriculation Number (a unique key), Age of student before admission, count of years before admission, program type (Full Time / Part Time), high school type (public/private), O.level exam, gender, and final grade. Selection of population element were by random across each level.

2.2 Method of Analysis

Three machine learning algorithms were used which include decision tree, Naïve Bayes, Support vector machine and one traditional classifier (the logistic regression) to classify students' final grade. The decision tree classifier is simply a flowchart diagram with the terminal nodes demonstrating classification decisions. Commencing with our dataset, measurement of the entropy was obtained to find a way to split the set until all the data belong to the same class. There are several approaches to decision trees like ID3, C4.5, CART and many more.

Naive Bayes uses the Bayes' Theorem in simple terms with an additional assumption that all predictors are independent. In other words, this classifier assumes that the presence of one particular feature in a class doesn't affect the presence of another one. It is similar to the Bayes' theorem however with more than one independent variable. The Bayes theorem given by (1) becomes (2) if there several independent variables.

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})} \quad \textbf{(1)}$$

$$\hat{y} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}} pv\ c\ (C_k) \prod_{i=1}^{n} p(x_i \mid C_k) \quad \textbf{(2)}$$

The Support Vector Machine (SVM) is a step further away from the support vector classifier which tries to classify a given data point based on two supporting edges of the distinct group (support vector) using either the maximal margin classifier (which is sensitive to outliers and do not leave room for misclassification) or the soft margin classifiers (which allows for bias at an advantage of lower variance). SVM uses either polynomial kernel which uses (3) to determine higher dimensional relationship from which classification is being made or the radial kernel which uses

(4). Where a and b are the support vectors (values at the edge of each clusters/groups), and value of r and d are best selected using cross-validation.

$$(a \times b + r)^d \quad (3)$$

$$e^{-\gamma(a-b)^2} \quad (4)$$

Finally the logistic regression is a logit transformation of the normal regression model which takes the form in (5) below:

$$\log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (5)$$

By simple algebra, we obtain p to be (6)

$$p = \frac{b^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{b^{\beta_0+\beta_1 x_1+\beta_2 x_2+1}} = \frac{1}{1+b^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}} = S_b(\beta_0 + \beta_1 x_1 + \beta_2 x_2) \quad (6)$$

Where S_b is the sigmoid function which is basis for finding the probability that Y = 1 (that is an event occurring). For the case where there are K levels of Y, the unordered log odd can be obtained using (7) and (8) for ordinal cases.

$$p(C_k \mid \phi) = y_k(\phi) = \frac{\exp(a_k)}{\sum_j \exp(a_j)} \quad (7)$$

where $\phi_k = [\phi_{k1}, .., \phi_{kM}]^{\mathrm{T}}, w_k = [w_{k1}, .., w_{kM}]^{\mathrm{T}}$

$a_k = w^T\phi + b$ is the "activation" $and\ b = biases$

$$\log \frac{P(Y \le j)}{P(Y > j)} = \text{logit}(P(Y \le j)) = \beta_{j0} - \eta_1 x_1^- \cdots - \eta_p x_p \quad (8)$$

Our model comprises 6 independent variables (Age at admission, Years before admission, Program type, High school type, 0'level, and Gender) against the ordered grade variable (Probation, Pass, Lower, Upper and Distinction). To ensure greater accuracy, algorithms were fed with varying training data of different ratios starting from 90% training and 10% testing to 50% training and 50% test. Records of varying performance measures were obtained on each occasion and a summarize result is obtained to compare between models. Comparison between algorithms and select the algorithm with the highest prediction accuracy was done. It is important to emphasize that a better alternative to our method is to use cross validation to juxtapose between performance of varying model like the study of Ramezan, Warner and Maxwell (2019) and Lope-del (2019) why they applied the population k-fold cross validation to selected the most precise model. However preference of the manual method is to monitor the trend of performance at each iteration.

3. Analysis and Presentation of Result

3.1 Data presentation

Table 1: Overview of data used for the study

SN	Matric Number	Age before admission	Years before admission	Program type	High school type	O'level Credits	Gender	grade
1	MS201500086	19	1	FT	NECO	5	M	Lower
2	MS201501470	20	3	FT	NECO	5	F	Lower
3	MS201501849	18	2	FT	NECO	7	F	Upper
4	MS201500568	18	1	FT	NECO	8	M	Lower
...	...	...	...	...	...	...	...	...
246	MS201602895	18	3	FT	NECO	8	F	Pass
247	MS201603549	18	1	FT	NECO	8	F	Lower
248	MS201602763	19	2	FT	WAEC	8	F	Pass
249	MS201603327	20	2	FT	NECO	8	F	Lower

Source: Federal Polytechnic Ede, Osun State

Table 1 shows the overview of our student academic data, a sample of 249 student record were obtained across the department of statistics (Federal Polytechnic Ede) for National diploma program. The target variable is the grade variable on 5 ordinal level (Probation, Pass, Lower, Upper and Distinction); other variables (except matric number and serial number) are the dependent variables used across all model fitted in this study.

3.2 Descriptive Analysis

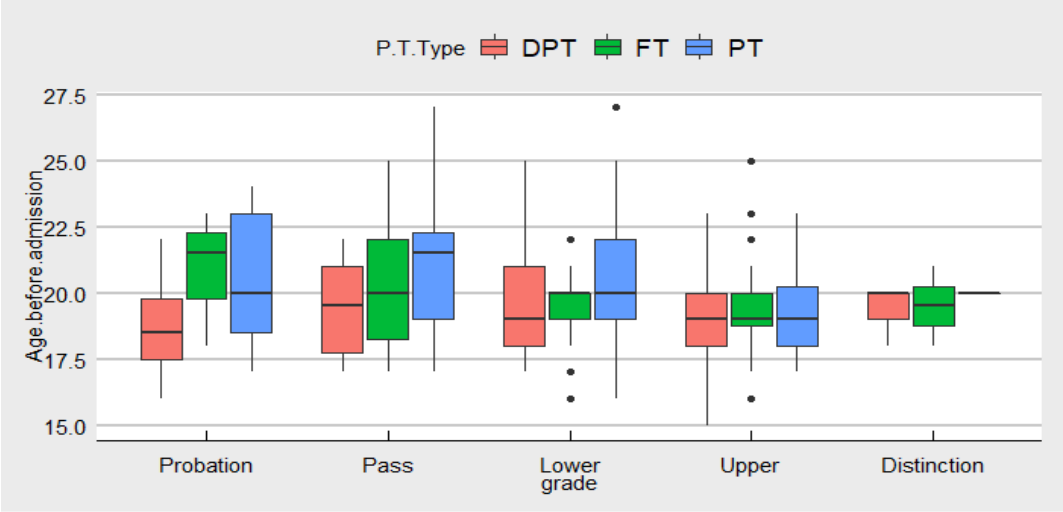
Table 2: Descriptive Analysis Table

Column	Min	Max	Mean	Std. deviation	Variance
Age before admission	15	27	19.96	2.1	4.41
Years before admission	1	10	2.41	1.27	1.62
O level	5	12	7.36	1.12	1.25

Source: Rstudio

A descriptive view into students characteristics indicated that minimum age of student before entrance into the Polytechnic is 15 years, average of 20 years and maximum age of 27 years. It took a minimum of 1 year to wait for admission but on the average 2 and at most 10 years after their secondary school activity.

Figure 1: Grade distribution by age



Source: Rstudio

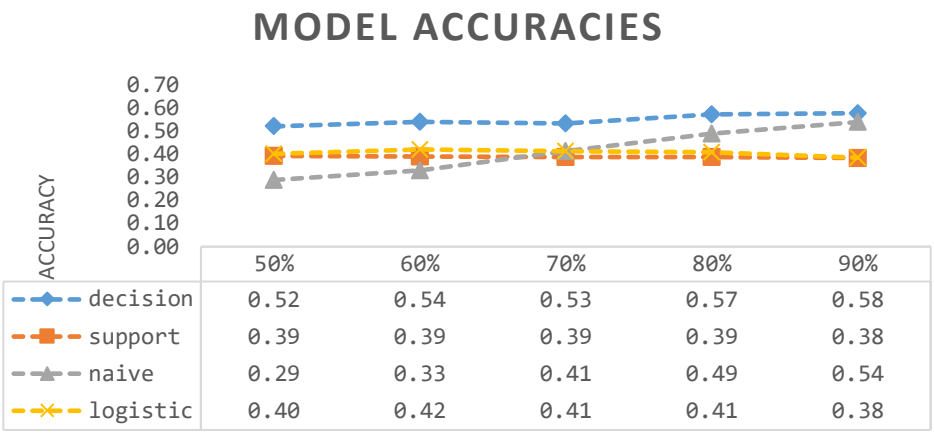
Judging from the grade distribution by age plot above, it is evident that people who studied full time and end up in probation have the maximum average age (over 21 years) while those that are admitted on daily part time program who ended up in probation are having the minimum average age (below 18 years). It is suggestible that age is a key factor that is been influenced by program type from the former result, however this claim also opposable by the latter result. This indicate that a more sophisticated inferential analysis is needed to make a good decision from this dataset.

3.3 Inferential Analysis

The analysis proceeded from the descriptive analysis to fitting four classification algorithms. The use of decision tree and support vector machine from the machine learning models was considered and the ordinal logistic regression and Naïve Bayes from the traditional data classifiers. The use of ordinal logistic regression is as a result of the target variable (grade) being on five ordered factor levels as previously started. R programming language was adopted to perform various training and testing iteration using the Caret package from the program and the accuracy of each model is shown in figure 2 below.

On all occasions, the result is similar to what is obtained in table 2 above. It is evident that decision tree outperforms other classification algorithm thus it is considered as the most appropriate algorithm for classifying our academic performance data. The conclusion from the analysis is similar to Sembiring et al., (2011) that machine learning algorithms are more efficient in classifying dataset than the traditional logistic regression analysis.

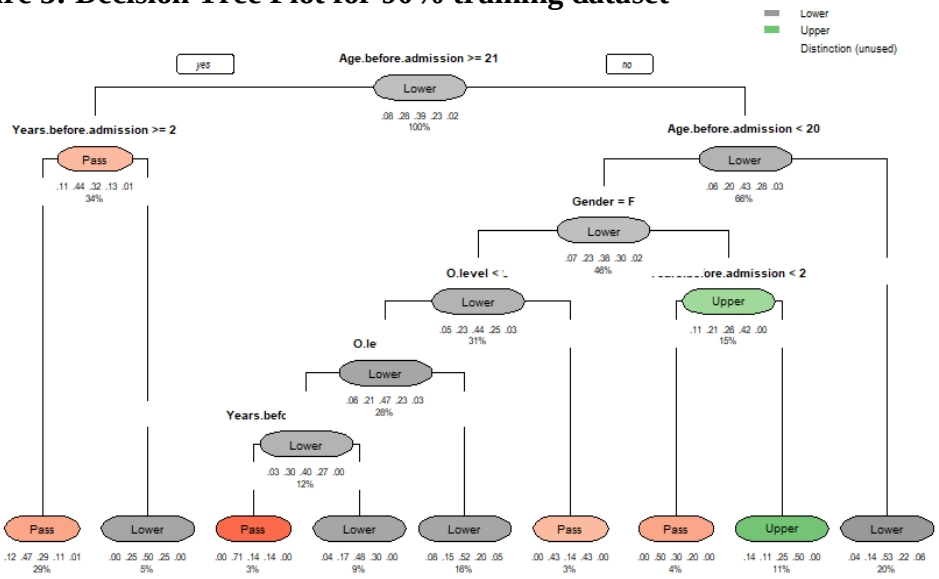
Figure 2: Model Accuracies



	50%	60%	70%	80%	90%
decision	0.52	0.54	0.53	0.57	0.58
support	0.39	0.39	0.39	0.39	0.38
naive	0.29	0.33	0.41	0.49	0.54
logistic	0.40	0.42	0.41	0.41	0.38

Source: Rstudio

Figure 3: Decision Tree Plot for 90% training dataset



Source: Rstudio

By observing the decision tree of 90% training data, it is evident that 21% of students who are older than a threshold of 21 years, lapsed more than 2 years before gaining admission, ending with a "PASS" grade while only 5% of those managed that reach the "LOWER" grade. On the other hand, 11% of male students who were younger than the threshold of 20 years and lapsed lesser than 2 years before gaining admission managed to achieve "UPPER CREDIT" while majority of female students end up either with a "PASS" or "LOWER CREDIT". The final conclusion of this study is that machine learning algorithms are more accurate in prediction making compared to traditional classifiers.

4. Conclusion and Recommendation

Based on the accuracy plot shown in figure 2, it is evident that all models improve as more training samples are fed, except for support vector machine which has a reduction in accuracy beyond 60%. It is also evident that decision algorithm performs best (though with a fair accuracy of 55% on average) compared to other models. This indicates that it is the best model that suit our data at 90% training data set. We as well noticed that Naive Bayes algorithms outperforms both support vector machine and logistic regression at 80% and 90% training dataset but also perform below and accuracy of 50% in training set below 50%. Conclusively we badge decision tree of 90% training set to be the best classification model in expense of other three models.

References

A Ramezan, C., A Warner, T., & E Maxwell, A. J. R. S. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *11*(2), 185.

Asif, M., Martiniano, H. F., Vicente, A. M., & Couto, F. M. J. P. o. (2018). Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. *13*(12), e0208626.

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. J. C. c. m. (2016). Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *44*(2), 368.

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Kweon, B.-S., Ellis, C. D., Lee, J., Jacobs, K. J. U. F., & Greening, U. (2017). The link between school environments and student academic performance. *23*, 35-43.

A Ramezan, C., A Warner, T., & E Maxwell, A. J. R. S. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *11*(2), 185.

Asif, M., Martiniano, H. F., Vicente, A. M., & Couto, F. M. J. P. o. (2018). Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. *13*(12), e0208626.

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. J. C. c. m. (2016). Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *44*(2), 368.

Sokkhey, P., & Okazaki, T. (2020). Hybrid machine learning algorithms for predicting academic performance. *Int. J. Adv. Comput. Sci. Appl, 11*(1).

Tomic, A., Tomic, I., Rosenberg-Hasson, Y., Dekker, C. L., Maecker, H. T., & Davis, M. M. J. T. J. o. I. (2019). SIMON, an automated machine learning system, reveals immune signatures of influenza vaccine responses. *203*(3), 749-759.

Xu, X., Wang, J., Peng, H., & Wu, R. J. C. i. H. B. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *98*, 166-173.

Zarlis, M., Hartama, D., & Wani, R. (2011). *Prediction of student academic performance by an application of data mining techniques.* Paper presented at the Proceedings of the International Conference on Management and Artificial Intelligence,(MAI, 11), IACSIT Press, Bali.

Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011). *Prediction of student academic performance by an application of data mining techniques.* Paper presented at the International Conference on Management and Artificial Intelligence IPEDR.

Sokkhey, P., & Okazaki, T. (2020). Hybrid machine learning algorithms for predicting academic performance. *Int. J. Adv. Comput. Sci. Appl, 11*(1).

Tomic, A., Tomic, I., Rosenberg-Hasson, Y., Dekker, C. L., Maecker, H. T., & Davis, M. M. J. T. J. o. I. (2019). SIMON, an automated machine learning system, reveals immune signatures of influenza vaccine responses. *203*(3), 749-759.

Wolff, S., O'Donncha, F., & Chen, B. J. J. o. M. S. (2020). Statistical and machine learning ensemble modelling to forecast sea surface temperature. *208*, 103347.

Xu, X., Wang, J., Peng, H., & Wu, R. J. C. i. H. B. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *98*, 166-173.

Zarlis, M., Hartama, D., & Wani, R. (2011). *Prediction of student academic performance by an application of data mining techniques.* Paper presented at the Proceedings of the International Conference on Management and Artificial Intelligence,(MAI, 11), IACSIT Press, Bali.