Data Mining Students' performance in a Higher Learning Environment

Ravneil Nand
School of Information Technology, Engineering, Mathematics and Physics
The University of the South Pacific
Suva, Fiji
nand_ra@usp.ac.fj

Ashneel Chand
College of Foundation Studies
The University of the South Pacific
Suva, Fiji
ashneel.chand@usp.ac.fj

Emmenual Reddy
School of Information Technology, Engineering, Mathematics and Physics
The University of the South Pacific
Suva, Fiji
reddy_e@usp.ac.fj

***Abstract*—Student performance in higher education has become one of the most widely studied area. While modelling students' performance, data plays a pivotal role in forecasting their performance and this is where the data mining applications are now becoming widely used. There are various factors which determine the student performance. In this study, eight attributes are used as inputs which are considered most influential in determining students' performance in the Pacific. Statistical analysis is done to find out which attribute has the highest influence to student performance. In this research, different algorithms are utilized for building the classification model, each of them using various classification techniques. The classification techniques used are Artificial Neural Network, Decision Tree, Decision Table, and Naïve Bayes. The dataset of 651 records used in this research is an imbalanced set, which is later transformed to balance set through under sampling. Neural Network is one of the classification techniques that has performed well on both, imbalanced and balanced datasets with the highest prediction accuracy of 96.8%. The analysis further shows that internal assessment has weak positive relationship with student performance while demographic data has no significant relationship.**

Keywords—Data mining, Classification algorithm, Artificial neural network, Decision tree, Decision table, Naïve Bayes

I. INTRODUCTION

One of the important aspect of a student's accomplishment at tertiary level education is their engagement in various activities such as assessments [1] [2]. These are either summative assessments or formative assessments. Ibrahim and Rusli in [3] alluded that the academic performance in their studies was measured by the student's cumulative grade point average (CGPA) upon graduation. Performance, therefore becomes a critical factor for student achievement; however, it can be influenced by a number of attributes. Inclusion of new technologies [4, 5, 6, 7] are also changing the tertiary level education landscape. Researchers have utilized a number of different tools and approaches to determine relationships and associations between different variables and factors such as student performance and online presence. One of the most influential approach of recent time has been data mining, which has become a very important method in understanding and solving educational issues at tertiary level [8] [9] and it aids in modelling students' performance. Different attributes such as the online engagement, student performance in tests and assignments play huge role in students' accomplishment [9].

Studies [3], [10], [11], [12], [13], [14] have used various data mining methods such as decision tree, and linear regression to predict students' performance and attrition rate. According to [15], the application of data mining techniques was seen in many areas such as business, telecommunication and banking as well as the education sector. The idea of having something unexpected or unknown builds curiosity that builds expectation and determination in developing or finding solutions through prediction models. Data mining has been used in different domains in developing methods for exploring the unique types of data [16] [17] [18]. It can be used in future for blockchain technologies [19].

Authors in [20] and [21] highlighted grade point average (GPA) and internal assessment as their main attributes to predict student performance. Shahiri [21] argued that these attributes are most widely used for predicting students' performance. The variables in internal assessment were assignments, quizzes, forums and short tests, lab work and attendance. There are other various factors such as gender, mode, assessment marks, participation in online activities which also need to be taken into consideration in order to make a more reliable and accurate forecasting of students' performance [13] [22]. Authors in [23], [12] and [24] highlighted that artificial neural network is most accurate when making predictions. Researchers in [11], [23] have used decision tree and linear regression when predicting students' performance.

The literature shows that data mining techniques such as the artificial neural network (ANN) and a combination of clustering and decision tree classification techniques for predicting and classifying students' academic performance seem to be mostly widely adopted by the many researchers [3], [10], [13], [14] and [25]. The rapid development and advancement of artificial intelligence and deep learning algorithms has provided another approach for intelligent classification and result prediction [26, 27, 24, 28, 29]. Moreover, as seen in Facial Recognition application research, more recently neural networks is a better classifier because it learns the features of data [30]. This gave motivation to see performance of neural network on small set data.

This research paper examines four common classification techniques namely decision table, decision tree, multilayer perceptron and Naïve Bayes to model students' performance. The motivation is to explore the importance of different variables of student performance in a higher education institute when applied to various classification algorithms. The outcome expected is that from the four algorithms, one would be the best one to use.

The two main contributions of this paper include:

1. Statistical methods for data mining are used to find the association between the variables.
2. Classification algorithms are used to find the best algorithm to model students' performance.

2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES) | 978-1-6654-2157-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/NILES53778.2021.9600504

978-1-6654-2157-7/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: University of the South Pacific. Downloaded on April 13,2022 at 05:46:43 UTC from IEEE Xplore. Restrictions apply.

The research paper is organized as follows: Section II is on methodology while section III gives the insight of the results obtained. Section IV presents the discussions, and Section V concludes the study with future recommendations.

II. METHODOLOGY

The objective of this research is met by a prediction model that was built to ascertain the passing or failing of students. The method used here is a quantitative approach as the data collected were students test marks and continuous assessments, and were analysed statistically. The classification models were used to analyse and to predict students' performance. Statistical analysis was carried out using R software. The data mining analysis was conducted using Weka through Decision table, Decision tree, Neural Network (NN) and Naïve Bayes classifiers.

A. Data Collection and Processing

Virtual learning platforms such as Moodle, allows higher education institutions to continuously and regularly monitor students' engagement in learning activities. The innovation allows tracking in real time. Moodle in the University of the South Pacific (USP) is managed by Information Technology Services (ITS). The data set collected has a total of 651 data entries where the clearance was sorted from respected departments. The log activities of these students were obtained from ITS. All of the log activities that was requested were time spent on course view, test marks, forum view, assignment marks and lecture capture view of students who were studying at 200 and 300 levels of their academic programme. The data that were collected concurrently included mode of study, gender and grades as shown in Table I. Grade of students were obtained from respective course facilitators where it was transformed binary as pass or fail.

The first step opted for data processing was initiating the process of data cleaning before normalisation. The real time data obtained were normalised using equation (1). The idea behind this was to minimise the data redundancy and improves data integrity [31].

$$\text{Normalized Value} = \frac{\left(x - \bar{x}\right)}{SD} \quad (1)$$

where, $\bar{x}$ is the mean and *SD* is the standard deviation.

TABLE I. ATTRIBUTES AND VALUES OF LOG ACTIVITIES

Attributes	Values	Description
Course level	200 level or 300 level	Level of course.
Gender	Male or Female	Student gender.
Test	Total Test Marks	Total Marks of students in test.
Assignment	Total Assignment Marks	Total Marks of students in Assignment.
Course view	Total Course views	Student Total Course views per semester.
Forum view	Total Forum views	Student Total Forum views per semester.
Lecture Capture View	Total Lecture Capture	Total Lecture Capture views per semester.
Mode	Blended or Online	Course offering mode.

In this research a confusion matrix was used to determine the model accuracy. Four algorithms that were used in this research include decision tree, decision table, neural network and Naïve Bayes algorithms. A question mark "?" was used where there was missing values of data. The data gathered was an unbalanced data and the model was tested for both balance and imbalance data. Under-sampling was used to create the balanced dataset. Balance data is where there were equal number of pass and fail, while imbalanced data had more pass (548) then compared to fail (103). The data arrangement was left random for both balanced and imbalance dataset. Different classification algorithms are applied during the performed research work, selected because they are likely to yield worthy results [32] [18]. Rules (decision tree), Trees (J48), Bayesian classifiers (Naïve Bayes) and Function (multilayer perceptron) are some dominant WEKA classifiers that have been used in this study. Random forest was not used in this research as it combines several decision trees in comparison to decision tree that is based on some decision. Thus, it is a long process, yet slow and not feasible in this research. Different test options namely; cross validation (10 fold and 20 fold), and percentage split (60% and 70%) were used in all above mentioned algorithms.

B. Statistical Analysis

Statistical analysis was carried in order to evaluate and compare the linear dependencies between two variables. The first step opted whereby the variables were classified into ordinal or nominal datasets. Based on the data type, best suited tests were conducted on different variables [9]. The common test carried out in this study was Pearson Test, which is a parametric correlation test where the variables are from normal distribution [33].

C. Performance Measurement

To evaluate and compare the performance of the classification models, classification accuracy, error rate and error type was used. A confusion matrix is used to calculate the same in this research. Fig. 1 shows the confusion matrix for two possible outcomes p (positive) and n (negative).

		Actual		Total
		p	*n*	
Predicted	*p′*	True Positive (TP)	False Positive (FP)	*P*
	n′	False Negative (FN)	True Negative (TN)	*N*
total		*P′*	*N′*	

Fig. 1. Confusion Matrix

The following equations (2) and (3) have been used to calculate classification accuracy, and true positive rate.

$$\text{Classification Accuracy} = \frac{\left(TP+TN\right)}{\left(TP+TN+FP+FN\right)} \quad (2)$$

$$\text{True Positive Rate} = \frac{TP}{\left(TP+FN\right)} \quad (3)$$

where, TP is true positive, TN is true negative while FP and FN are false positive and negative, respectively.

Authorized licensed use limited to: University of the South Pacific. Downloaded on April 13,2022 at 05:46:43 UTC from IEEE Xplore. Restrictions apply.

III. RESULTS

A. Association

In this sub-section, associations between variables were investigated using the Pearson's correlation test as seen in tables II-III.

The linear associations between the Test, Assignment (Assign.), Lecture View (Lect_View) and the Grades were investigated using the Pearson's correlation test as shown in Table II. It can be seen there is a weak positive linear relationship between Test and Grades, and Assignment and Grades since R values are in the range of 0.3 and since $p<0.05$, the correlation coefficient are statistically significant. As for Lecture View it is not statistically significant since R value is near to zero.

TABLE II. PEARSON'S CORRELATION TEST ON THREE VARIABLES

		Test	Assign.	Lect_View	Grade
Test	***Correlation***	1.000	-0.21	0.22	0.35
	Sig.	-	3.08	1.22E-08	**1.42E-20**
Assign.	***Correlation***	-0.21	1.000	-0.11	0.35
	Sig.	3.08E-20	-	0.004	**3.82E-20**
Lect_View	***Correlation***	0.22	-0.11	1.00	0.04
	Sig.	1.22E-08	0.004	-	0.27
Grades	***Correlation***	0.35	0.35	0.04	1.0
	Sig.	1.42E-20	3.82E-20	0.27	-

TABLE III. PEARSON'S CORRELATION TEST ON FIVE VARIABLES

		Grade
CourseView	***Correlation***	**-0.113**
	Sig. (2-tailed)	0.0040
ForumView	***Correlation***	**-0.128**
	Sig. (2-tailed)	0.0010
Gender	***Correlation***	**-0.110**
	Sig. (2-tailed)	0.0048
Mode	***Correlation***	**-0.141**
	Sig. (2-tailed)	0.00030
Levels	***Correlation***	**0.076**
	Sig. (2-tailed)	0.0516

Table III shows the Pearson correlation results of CourseView, ForumView, Gender, Mode, Levels with Grades. CourseView, ForumView, Gender and Mode have a weak negative correlation with grade and since $p<0.05$, its statistically significant. If the p-value is small, there is a statistically significant correlation. As for "Level" it is not as its R value is closer to zero indicating zero correlation.

B. Balanced data

This sub-section shows results of classifiers on the balanced data.

Tables IV – V shows the accuracy (Acc.) and True Positive Rate (TPR) of various classifiers for balanced data. Table IV shows the results of percentage split of different classifiers where P and F were arranged in order (Balanced Data). The two models which produced best results are the Decision tree and the Artificial Neural Network (ANN). Both models produced more than 96% accuracy.

TABLE IV. PERCENTAGE SPLIT OF DIFFERENT CLASSIFIERS FOR BALANCED DATA

Class (%)	Decision Table		Decision Tree		Neural Network		Naïve Bayes	
	Acc.	*TPR*	*Acc.*	*TPR*	*Acc.*	*TPR*	*Acc.*	*TPR*
60	92.7	90.4	**96.3**	**94.1**	95.1	94	93.4	93.9
70	95.2	91.4	95.2	91.4	**96.8**	**94.1**	95.2	91.4

Table V shows the results of cross validation for different classifiers where P and F were arranged in order. Again it was seen that Decision tree and ANN had better results than decision table and naïve Bayes models. Fig. 2 shows classifiers accuracy comparison for Balanced Data.

TABLE V. CROSS VALIDATION OF DIFFERENT CLASSIFIERS FOR BALANCED DATA

Folds	Decision Table		Decision Tree		Neural Network		Naïve Bayes	
	Acc.	TPR	Acc.	TPR	Acc.	TPR	Acc.	TPR
10	90.3	85.5	92.7	88.6	**93.7**	**90.2**	92.3	87.8
20	90.3	85.5	**92.7**	**88.6**	**92.7**	**87.8**	92.3	87.8

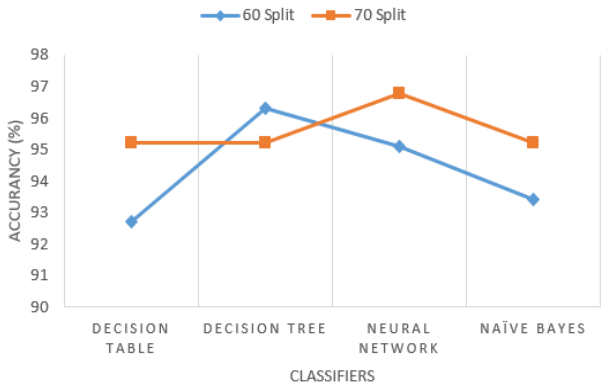

Fig. 2. Classifiers Accuracy Comparison for Balanced Data

C. Imbalanced data

In this sub-section, the performance of various classification algorithms is discussed on imbalanced dataset.

Tables VI – VII show the accuracy (Acc.) and True Positive Rate (TPR) of various classifiers for imbalanced data. Bold values indicate the best results obtained. Table VI shows the results of percentage split of different classifiers where P and F was not arranged in order. It can be seen that the Decision table is one of the algorithms with best results. In both percentage (%) split, the performance is best for Decision Table when compared to other three methods. Table VII shows the results of cross validation for different classifiers where P and F was not arranged in order. This time around, Decision Table and ANN had the best results in comparison to other two methods. The accuracy for both these methods is 89.4%.

TABLE VI. PERCENTAGE SPLIT OF DIFFERENT CLASSIFIERS FOR IMBALANCED DATA

Class (%)	Decision Table		Decision Tree		Neural Network		Naïve Bayes	
	Acc.	TPR	Acc.	TPR	Acc.	TPR	Acc.	TPR
60	**88.1**	**98.1**	83.8	93.4	85	97.6	81.9	90.2
70	**87.2**	**97.5**	85.6	95.6	86.2	96.9	81.5	91.9

TABLE VII. CROSS VALIDATION OF DIFFERENT CLASSIFIERS FOR IMBALANCED DATA

Folds	Decision Table		Decision Tree		Neural Network		Naïve Bayes	
	Acc.	TPR	Acc.	TPR	Acc.	TPR	Acc.	TPR
10	**89.4**	**90**	88.8	90.9	**89.4**	**91**	86.5	90.4

Authorized licensed use limited to: University of the South Pacific. Downloaded on April 13,2022 at 05:46:43 UTC from IEEE Xplore. Restrictions apply.

20	**89.4**	**90.1**	88	90.1	**89.4**	**90.8**	86	90.2

Fig. 3 shows classifiers accuracy comparison for imbalanced Data.

Fig. 3. Classifiers' Accuracy Comparison for Imbalanced Data

D. Overall

This sub-section shows the overall performance of various classification algorithms.

Table VIII shows the ranking of accuracy of various classifiers for balanced and imbalanced dataset. The classifiers are represented by letters; decision table (T), decision tree (T), neural network (N) and Naïve Bayes (B). The best results are highlighted in bold.

TABLE VIII. RANKING OF DIFFERENT CLASSIFIERS.

Method	Balanced Data				Imbalanced Data			
	D	T	N	B	D	T	N	B
60 % Split Data	4	1	2	3	1	3	2	4
70 % Split Data	2	2	1	2	1	3	2	4
10 Folds	4	2	1	3	1	3	1	4
20 Folds	4	1	1	3	1	3	1	4
Total Rank	14	6	5	11	4	12	6	16
Mean Rank	4	**1.5**	**1.3**	2.8	**1**	3	**1.5**	4

IV. DISCUSSION

In this section, the discussion is based on the individual performance of the algorithms.

The two methods that have outperformed in imbalanced dataset are Neural Network and Decision Table while Neural Network got best results for balanced datasets. It can be seen that in balanced data, Decision Table has the worst results while in imbalanced data it is Naïve Bayes classifier. A classification technique which has produced best results in both datasets is Neural Network. While its performance is second to Decision Table in imbalanced dataset, it has a consistent performance.

Tables IV-VII show also the highest prediction accuracy of each method with respect to the dataset. The highest prediction accuracy of Neural Network is (96.8%) followed by Decision Tree by (96.3%). Next, Naïve Bayes and Decision Table gave the same accuracy, which is (95.2%). All the methods in this research have highest accuracy recorded in balanced dataset. According to literature, under sampling can sometimes lose some important attributes [34] but in this case, it has proved to be beneficial. The data is not noisy, therefore, classifiers work well with balanced dataset.

Tables VIII shows results based on ranking. It can be seen that Neural Network has the best ranks in balanced data while decision tree has in imbalanced data. For the balanced data, Neural Network is able to achieve rank number 1 in 3 out of 4 (75%) instances while decision tree it is in 2 out of 4 (50%) cases while in imbalanced data, Decision Table is able to achieve number 1 rank in 4 out of 4 (100%) cases while Neural Network it is 2 out of 4 (50%) cases. Mean rank for neural network is consistent in both cases. Both figures (Fig. 2 and Fig. 3) show different classifier accuracy rate on two datasets. Neural network can be seen with a consistent performance.

The result of prediction accuracy is highly dependent on the attributes or features that were used during the initial prediction process. Neural Network method gave the highest prediction accuracy because of the influence from main attributes. The attributes Test, Assignment, CourseView, ForumView, LCaptureView, Gender, Mode and Levels have all played a very important role prediction accuracy; as it was found that there is a weak correlation between the variables. According to the statistical analysis seen in Tables II-III, the two most influential attributes are Test and Assignment.

One of the advantages of Neural Network is the ability to capture nonlinear relationships easily. It is also referred as adaptive system due to its ability to readily update the historical data like a human brain. Therefore, the model always functions beyond the knowledge base. In addition, the strength of neural network is the ability to learn from a limited dataset.

V. CONCLUSION

The prediction of student academic performance has drawn considerable attention in Higher Education environments. The most popular data mining technique in predicting student's performance is the classification model. The classification algorithms; Decision Tree, Artificial Neural Networks (ANN), Naive Bayes, and Decision Table are used in this research to predict student performance through use of different attributes.

One method that has shown good performance in balanced and imbalanced dataset is Neural Network with the highest prediction accuracy of 96.8%, the increase due to internal assessments. The other variables have underplayed in this research. Decision Tree followed with a 96.3% accuracy. Neural Network and Decision Tree are the two methods highly used by researchers for predicting student's performance and these have also performed well on the datasets used in this research.

This research has allowed us to look into other variables that can influence the performance that is meta-analysis. It will help the educational system to monitor the students' performance in a better way where important features need

Authorized licensed use limited to: University of the South Pacific. Downloaded on April 13,2022 at 05:46:43 UTC from IEEE Xplore. Restrictions apply.

special attention. Socio-economic factors of the students are worth analysing in the extension of this research.

REFERENCES

[1] R. K. Orchard, "Multiple attempts for online assessments in an operations management course: An exploration," Journal of Education for Business, vol. 91, no. 8, pp. 427-433, 2016.

[2] E. Faulconer, J. C. Griffith and H. Frank, "If at first you do not succeed: student behavior when provided feedforward with multiple trials for online summative assessments," Teaching in Higher Education, pp. 1-16, 2019.

[3] Z. Ibrahim and D. Rusli, "Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression," in 21st Annual SAS Malaysia Forum, 5th September, 2007.

[4] P. Reddy, E. Reddy, V. Chand, S. Paea and A. Prasad, Assistive technologies: saviour of Mathematics in higher education, vol. 6, Frontiers, 2020, p. 75.

[5] S. H. Raza and E. Reddy, "Intentionality and Players of Effective Online Courses in Mathematics," Frontiers in Applied Mathematicsand Statistics, vol. 7, p. 612327, 2021.

[6] E. Reddy and B. Sharma, "Mobile learning perception and attitude of secondary school students in the Pacific Islands," in 22nd Pacific Asia Conference on Information Systems, Yokohama, 2018.

[7] P. Reddy, K. Chaudhary, B. Sharma and R. Chand, "The two perfect scorers for technology acceptance," Education and Information Technologies, vol. 26, no. 2, pp. 1505-1526, 2021.

[8] M. Agaoglu, "Predicting instructor performance using data mining techniques in higher education," IEEE Access, vol. 4, pp. 2379-2387, 2016.

[9] B. Sharma, R. Nand, M. Naseem and E. V. Reddy, "Effectiveness of online presence in a blended higher learning environment in the Pacific," Studies in Higher Education, pp. 1-19, 2019.

[10] P. Baepler and C. J. Murdoch, "Academic analytics and data mining in higher education," International Journal for the Scholarship of Teaching and Learning, vol. 4, no. 2, pp. 1-9, 2010.

[11] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," arXiv preprint arXiv:1201.3417, 2012.

[12] S. Borkar and K. Rajeswari, "Attributes selection for predicting students' academic performance using education data mining and artificial neural network," International Journal of Computer Applications, vol. 86, no. 10, 2014.

[13] B. Guo, R. Zhang, G. Xu, C. Shi and L. Yang, "Predicting students performance in educational data mining," in 2015 International Symposium on Educational Technology (ISET), 2015.

[14] S. Roy and A. Garg, "Analyzing performance of students by using data mining techniques a literature survey," in 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017.

[15] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang and H. Y. Seong, "Predicting NDUM student's academic performance using data mining techniques," in 2009 Second International Conference on Computer and Electrical Engineering, 2009.

[16] M. Naseem, K. Chaudhary, B. Sharma and A. L. Goel, "Using Ensemble Decision Tree Model to Predict Student Dropout in Computing Science," in IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Melbourne, Australia, 2019.

[17] R. Nand, A. Chand and M. Naseem, Analyzing students' online presence in undergraduate courses using Clustering, IEEE, 2020, pp. 1-6.

[18] M. H. Aung, P. T. Seluka, J. T. R. Fuata, M. J. Tikoisuva, M. S. Cabealawa and R. Nand, Random Forest Classifier for Detecting Credit Card Fraud based on Performance Metrics, IEEE, 2020, pp. 1-6.

[19] K. Chaudhary, V. Chand and A. Feknker, "Double-Spending Analysis of Bitcoin," in Pacific Asia Conference on Information Systems, Dubai, 2020.

[20] A. C. Hachey, C. Wladis and K. Conway, "Prior online course experience and GPA as predictors of subsequent online STEM course outcomes," The Internet and Higher Education, vol. 25, pp. 11-17, 2015.

[21] A. Shahiri, "M., Husain, W., Rashid, N," A., A Review on Predicting Students Performance using Data Mining Techniques. Procedia Computer Science, vol. 72, pp. 414-422, 2015.

[22] R. S. Baragash and H. Al-Samarraie, "Blended learning: Investigating the influence of engagement in multiple learning delivery modes on students' performance," Telematics and Informatics, vol. 35, no. 7, pp. 2082-2098, 2018.

[23] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," Computers & Education, vol. 61, pp. 133-145, 2013.

[24] B. S. Kalyani, D. Harisha, V. RamyaKrishna and S. Manne, "Evaluation of Students Performance Using Neural Networks," in International Conference on Intelligent Computing, Information and Control Systems, 2019.

[25] R. Nand, E. Reddy and M. Naseem, "Neuron-network level problem decomposition method for cooperative coevolution of recurrent networks for time series prediction," in Lecture Notes in Computer Science, Springer Verlag, 2016, pp. 38-48.

[26] W. W. T. Fok, Y. S. He, H. H. A. Yeung, K. Y. Law, K. H. Cheung, Y. Y. Ai and P. Ho, "Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine," in 2018 4th International Conference on Information Management (ICIM), 2018.

[27] M. Zaffar, S. Iskander and M. A. Hashmani, "A study of feature selection algorithms for predicting students academic performance," Int. J. Adv. Comput. Sci. Appl, vol. 9, no. 5, pp. 541-549, 2018.

[28] K. Chaudhary and D. X, "P2P-netpay: An off-line micro-payment system for content sharing in P2P-networks," Journal of Emerging Technologies in Web Intelligence, vol. 1, no. 1, pp. 46-54, 2009.

[29] X. Dai, K. Chaudhary and J. Grundy, "Comparing and contrasting micro-payment models for content sharing in P2P networks," in 3rd IEEE International Conference on Signal Image Technologies and Internet Based Systems, Shanghai, 2007.

[30] S. Paul and S. K. Acharya, A Comparative Study on Facial Recognition Algorithms, 2020.

[31] A. Chand and R. Nand, "Rainfall prediction using Artificial Neural Network in the South Pacific region," in 2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2019.

[32] D. Kabakchieva, Student performance prediction by using data mining classification algorithms, vol. 1, 2012, pp. 686-690.

[33] S. Arndt, C. Turvey and N. C. Andreasen, "Correlating and predicting psychiatric symptom ratings: Spearmans r versus Kendalls tau correlation," Journal of psychiatric research, vol. 33, no. 2, pp. 97-104, 1999.

[34] X.-Y. Liu, J. Wu and Z.-H. Zhou, Exploratory undersampling for class-imbalance learning, vol. 39, IEEE, 2008, pp. 539-550.

Authorized licensed use limited to: University of the South Pacific. Downloaded on April 13,2022 at 05:46:43 UTC from IEEE Xplore. Restrictions apply.