Machine Learning for Strategic Decision Making during COVID-19 at Higher Education Institutes

Amjed Sid Ahmed Mohamed Sid Ahmed
Department of Computing
Global College of Engineering and Technology
Muscat, Oman
0000-0003-1236-1893

Mazhar Hussain Malik
Department of Computing
Global College of Engineering and Technology
Muscat, Oman
0000-0001-8890-1310

***Abstract*—Machine learning is becoming driving force for strategic decision making in higher educational institutions and it calls for cooperation between stakeholders and the use of efficient computation methods. Contrariwise, making decisions might consume much time, if there is no use of data and computational methods during the process of decision making.The utilization of machine learning is essential when coming up with an ultimate analysis of data and decision making. Besides, the technology which is under artificial intelligence could facilitates incredible output for educational institutes when it came to decision making. This paper analyses the output generated using machine learning algorithms that help in prediction of no detriment policy applicability rate in the case of e-learning during COVID-19. The study investigates the performance of machine learning algorithms for strategic decision making in the higher educational institutes, Global College of Engineering and Technology in particular, whether no detriment policy will be applicable for a particular student based on students performance before COVID-19. The study shown that Random Forest machine learning algorithm performance is higher as compare to Support Vector Machine, Decision Tree and Navie Bayes.**

***Index Terms*—COVID-19, Machine Learning Algorithms, No Detriment Policy, Strategic Decisions,**

I. INTRODUCTION

Meeting stakeholders needs and expectations require to utilize ICT (Information Communications Technologies) technology to establish a efficient generation of records and storage. Besides, institutions need to develop more advanced and high-quality services systems. In general, most of universities have focused on technology systems that enhance online learning, decision making, quality assurance, and academic research.

A. AI in Education Sector

Artificial intelligence in Education fields has been researching to know how to apply AI technologies to establish a suitable platform for students. Majority of these researches focused on coming up with systems that follow instructions provided by the user [1]. Their main goal was utilizing intelligent pedagogical agents (IPA) that would facilitate the learning process. These systems would have the capability to predict and evaluate a particular student's behavior when undertaking educational activities. Moreover, the designs would help students have appropriate feedback on complex issues giving possible solutions [2]. To make these technologies efficient, IPA systems must-have capabilities of capturing and analyzing data based on the three main teaching instruments. Teaching components include the domain model, pedagogical model, and the particular students' models. Besides, they were the basics of instructional model architecture founded by Intelligent Tutoring Systems (ITS) [3]. Basin the argument on the student's model, systems that do not utilize AI technology, for instance, test-branch devices, gave students feedback. Besides, they worked under codes that they could connect student's responses to provide suitable outcomes [4]. In contrast, ITS research focuses on coming up with highly detective systems that have precise output. The designs are set to perform under high pedagogical processes with characteristics that can give instant feedback from users. Within these type of systems, users' instructions will consist of an individual's intelligence and capabilities, meta-cognitive skills, and emotions. ITS research has established systems that are suitable for an individual's output. Programs have enabled students to code, work on geometry, physics, and algebra, among other domains. Besides, these systems have played a vital role in improving students' performance compared to class or group discussions [5]–[8]. Moreover, ITS technology applications have developed a more reliable system that has helped curb teacher's shortage and overcoming the old teaching methods. Furthermore, it has facilitated students' influx in schools [9], [10]. All in all, it has provided a convenient platform that calls for investments to foster this technology. It will continue affecting the performance of humans positively and reduce expenses.

B. Strategic Planning and Decision Making

Strategic planning could be defines as a company's scheme that shows its position, strengths, and weaknesses and prioritizes profitable activities [11]. A strategy is a plan that determines and brings together primary objectives, set rules, and sequential processes [12]. Long term planning focuses on the future demands of the business and how to cope with future situations. Besides, strategic planning is plotting and maintaining profitable objectives that will help the organization to withstand market variations [13]. According to [14]strategic planning is a sequential process of establishing risky decisions and analyzing possible future outcomes. The process requires the company to have a proper organization of capabilities to facilitate the process. Moreover, measuring the result of the decisions against their expectations. Besides, the organization has to come up with adequately organized responses. From a general perspective, long term planning can be defined as an institutional process of determining their goals and making decisions on how to allocate resources to meet its objectives. The country's prestige in the world depends on education [15]. Educational responsibilities, especially by the tertiary institutions, determine various careers such as technicians and specialists. These people play a vital role in boosting and building a country's economy. Therefore, it calls for better decision-making strategies in institutions since they directly influence people and their environment. Besides, it has an impact on their political, social, and economic backgrounds. The entire academic fraternity is directly affected by the complex decisions made by directors of these institutions. To aid the decision-making processes, managerial officials depend on data tools. The advanced decision-making processes utilize extracted data for computations and analyzing various connections. Besides, the information tools contribute to learners and tutors [16]. In this era, expert system codes have become more accurate than other statistical systems [17]. Artificial Neural Networks (ANN) and Support Vector Machines(SVM) have aided both practical and theoretical research due to the accuracy when making decisions [18]. Besides, both methods have different higher precision capabilities for ultimate performance. The main contribution of the study is to investigate the impact of machine learning algorithms on no detriment policy during COVID-19. The study verify the accuracy of machine learning algorithms using pandemic time data of students performance at Global College of Engineering and Technology as case study.

II. Literature Review

Since 1970, there has been a remarkable evolution and development of programs that work per decision support systems (DSS) [19]–[23]. Over 40 years, the systems have had a better performance on computations of data. Besides, there has been a faster functionality of the algorithms during data processing. The progress has contributed to the development of human interface programs that handle vast and complex data. Currently, there has been the advancement of the DSS to an educational support system. The new system contributes to the education sector. There has been a business intelligence system that analyses and develops knowledge extracted from online studies over the internet. Moreover, data mining carries out analysis of data obtained from online data sources. Besides, the data mining technology utilizes high-speed processors. Enhancing educational fields called for the development of educational mining systems (EDM). To establish a better and clear understanding of studying and the essence of sciences cognitive processes, learning analytics were developed. Lastly, the machine learning systems raised with capability of doing data analysis on extracted data using complex algorithms. Considering the education field, machine learning has helped in identifying the performance of institutions. It has helped in the evaluation of quality teaching, providing a response, giving possible predictions, choosing of best products and tests [24]–[28]. Besides, there has been a better strategy for each course and evaluation of outcomes from different educational methods [29]. For instance, institutions use artificial neural networks and support vector machine algorithms. Besides, further researches shows the use of linear, logistic regression and random forest algorithms. Data classification can be done using random forest, logic regression and artificial neural network algorithms.. [30] investigating possibility of detecting the rates of school dropouts. They used populations where they branched the groups into two, "success" and "failure." They subdivided the success' groups further to come up with two other unknown divisions. Based on that, an experiment was set to determine the number of graduates on a particular split data. The main aim of carrying out the practical activity was to measure these machine learning algorithms' accuracy. In conclusion, most academic research focuses on tutors, learners, and particular school administrators [31]. Most of authors of current researches focused on student's performance when undertaking their studies. Moreover, the research does not factor decision processes made by managers as one of the analysis's essential tools. Generally speaking and up to authors best knowledge, machine learning fields have not yet come up with ways to handle the complex decision making processes by directors in tertiary institutions.

III. Machine Learning Algorithms for Decision Making in Higher Education

A. Support Vector Machine

Support Vector Machine (SVM) is machine learning algorithm which is used in supervised learning. SVM can be used for both classification and regression but mostly is used for classifications. In SVM algorithm, n-dimensional space is use to plot data in points; while n denote the total number of the present features and value of the feature points to particular coordinates [32]. SVM Polynomial Kernel function used in the support vector (SVMs) and similar kernelized based models. The algorithm provides vectors similarity in feature space which enables learning of non-linear models [33].

The polynomial kernel is defined in Eq.1:

$$Kn(a,b) = (a^T b + c)^z \tag{1}$$

As kernel Kn corresponding mapping as shown in Eq.2

$$K(a,b) = (\sum(i=1)^n a_i\ b_i + c)^2 = \sum_{i=1}^{n} (a_i^2)\ (b_i^2) \sum_{i=2}^{n} \sum_{j=1}^{i-1} (\sqrt{2a_i}a_j)\ (\sqrt{2b_i}b_j) + \sum_{i=1}^{n} (\sqrt{2ca_i})\ (\sqrt{2cb_i}) + c^2 \tag{2}$$

From this, it follows that the feature map is given in Eq.3

$$\varphi(a) = (a_n^2, \ldots,\ a_{1,}^2\ \ \sqrt{2}a_n a_{n-1}, \ldots, \sqrt{2}a_n a_{1,}, \sqrt{2}a_{n-1}a_{n-2}, \ldots, \sqrt{2}a_{n-1}a_1, \ldots, \sqrt{2}a_2 a_1, \sqrt{2}\mathrm{c}a_n, \ldots, \sqrt{2}\mathrm{c}a_1, c) \quad (3)$$

B. Random Forest

Random forest algorithm creates decision trees on data used for samples and algorithm use voting to get the best possible solution [34], [35].

C. Decision Tree

Decision tree also belongs to the supervised learning algorithms family. Decision tree can also use to solve both regression and classification problem and normally used for classifications [36], [37]. The basic principle behind the decision tree algorithm is to create a training model which helps to predict the values of required variables and class by implementing the learning decision rules inferred by the prior data mean during data training. Decision tree is easy to understand as compare to other algorithms and solve problem by using the tree representation. Where each node of the tree corresponds to an attribute and each lead node corresponds to a class label [38], [39].

Algorithm 1: Decision Tree

Result: Pseudocode for decision tree
1) Place the best attribute of the dataset at the root of the tree.
2) Split the training set into subsets. Subsets should be made in such a way that each subset contains data for an attribute.
3) Repeat step 1 and step 2 to find the leaf nodes in all tree branches.

D. Naive Bayes

Naive Bayes a probabilistic graphical model that represents knowledge dealing with random variables [40]. In model, random variable are denoted by the nodes, and the edges between the variables represent the conditional dependencies. Conditional dependencies are calculated by statistical probabilistic theories and computational methods. Naive Bayes depends on the Bayesian theorem and use conditional probability which is a powerful algorithm for predictive modelling. Additionally, the Naive Bayes classifier works quite well concerning real-world situations such as spam filtering. Bayes theorem is stated as mathematically as in Eq.4:

$$P\left(X|Y\right) = \frac{P\left(Y|X\right)P\left(X\right)}{P\left(Y\right)} \quad (4)$$

Where X and Y are events and $P(Y)?0$ [41].

- In this case, we are trying to find the probability of X, given that the event Y is true.
- P(X) is the priori of X (the prior probability, i.e. Probability of event before evidence is seen).
- P(X—Y) is a posteriori probability of Y

Now, with regards to the dataset, We can apply Bayes' theorem as per Eq.5:

$$P\left(y|x\right) = \frac{P\left(x|y\right)P(y)}{P(x)} \quad (5)$$

Where y denotes the class variable and x denotes the dependent feature of vector (of size n) as written in Eq.6:

$$X = (x_1,\ x_2,\ x_3,\ \ldots..,\ x_n) \quad (6)$$

Naive assumption If two events X and Y are independent, then the probability will be written as Eq.7,8.

$$P(X,Y) = P(X)P(Y) \quad (7)$$

Hence:

$$P\left(y|x_1,\ \ldots, x_n\right) = \frac{P\left(x_1|y\right)P\left(x_2|y\right)\ldots P\left(x_n|y\right)P(y)}{P\left(x_1\right)P\left(x_2\right)\ldots P(x_n)} \quad (8)$$

This can be written as Eq.9,10.

$$P\left(y|x_1,\ \ldots, x_n\right) = \frac{P\left(y\right)\prod_{i=1}^{n} P\left(x_i|y\right)}{P\left(x_1\right)P\left(x_2\right)\ldots P(x_n)} \quad (9)$$

Hence:

$$P\left(y|x_1,\ \ldots, x_n\right)\ \alpha\ P\left(y\right)\prod_{i=1}^{n} P\left(x_i|y\right) \quad (10)$$

Now, there should create a classifier model, in this case, we should know the probability of given set of inputs for class variable y and need to select the output which has maximum probability. The equation can be written as Eq.11.

$$y = argmax_y\ P\left(y\right)\prod_{i=1}^{n} P\left(x_i|y\right) \quad (11)$$

So, finally, we are left with the task of calculating P(y) and $P(xi|y)$ and $P(y)$ is also called class probability and $P(xi|y)$ is called conditional probability.

IV. Proposed Model

The purpose of the study is to compare the performance of Machine learning algorithms in predicting no-detriment rate that will support decision making at the strategic level to address the issues of students results during the COVID-19. Thus, we classify the student's performance to predict whether no detriment policy will be applicable to student(s) in the case of low performance as compare to before COVID-19 performance. The confusion matrix is shown in Table.1. From the confusion matrix data, The precision rate of the no

TABLE I
CONFUSION MATRIX

	No detriment applicable	No detriment not applicable	Total
No detriment applicable	w	x	$w+x$
No detriment not applicable	y	z	$y+z$
Total	$w+y$	$x+z$	N

detriment applicable class = $w/(w+z)$ the precision rate of the no detriment not applicable class = $x/(x+y)$. The recall rate of the no detriment applicable class = $w/(w+y)$ and the recall rate of the no detriment not applicable class $x/(x+z)$ and the overall accuracy $(w+x)/(w+x+y+z)$. To perform the extraction, transformation, and loading, we use WEKA which free software for data analytics. A record of 1020 students of Global College of Engineering and Technology (GCET), Oman is obtained out of that 85% data is used for machine learning purpose, while 15% data which is about 150 students record is used to predict whether students academic result will be supported by the no detriment policy? The list of variables and measurements are shown in Table.II.

TABLE II
VARIABLE AND MEASUREMENTS

Module Enrolled in Current Semester	Quantity
Modules passing Rate each semester	Quantity
Modules fail rate each semester	Quantity
Overall average percentage	Quantity
Highest average percentage in a semester	Quantity
Lowest average percentage in a semester	Quantity
Modules passing rate overall	Quantity
Modules fail rate overall	Quantity
Median percentage	Quantity
Percentage between the median and higher percentage	Quantity
The square root of the average	Quantity

V. RESULTS DISCUSSION

Table. III and Fig.1. shows that performance of all four algorithms which are used on the given dataset. The overall performance of Random Forest is very high and contributes to 0.99. decision tree rules able to get the correlation coefficient value of 0.86, while Gaussian Processes are at 0.723 and naive bayes contributes to 0.717 of correlation coefficient. MAE value of random forest is just 0.005 which is also very good and low, in the case of decision tree rules the value is 188.43 which is second best option as compare to random forest while SVM and naive bayes is contributing to 301.34 and 298.65 of MAE which shows less efficiency of the technique, same is the case of root mean square error for tree algorithms are low and it is also worth to mention that RAE and RRAE of greedy is lower than random forest methods and it is about 9% and 17% and while SVM and naive bayes are performing worst.

The Table.Table. IV is showing the accuracy of the algorithm by distributed into different classes. TP rate and FP Rates are rates of True Positive and False Positive of the tree classification algorithm of each class. No detriment applicable TP rate is 1.00 which is maximum achieve able rate, while

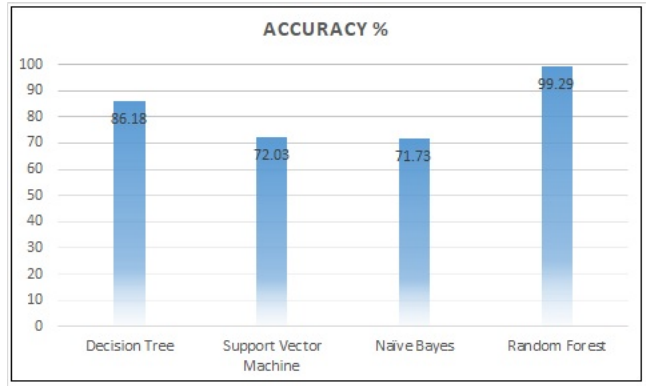

Fig. 1. Algorithms Accuracy Analysis

in the case of No detriment not applicable TP rate is 0.983 and total weight of TP rate is 0.999 and overall FP rate is just 0.013. Precision is related with the overall accuracy which is calculated as $TP/(TP+FP)$ which is 0.999 in our case which support algorithm in the way the outcomes are achieved correctly. The value of Recall is calculated as = $TP/(TP+FN)$ where FN is false negative which also showing reflection of True Positives outcomes contributes to 0.999 overall. Also it is notice that No detriment not applicable contribute to 0.991 Recall this is due to nature of the protocols which still contribute to maximum Recall. F-Measure is combined measure for precision and recall calculated as $2 * Precision * Recall/(Precision + Recall)$ contribute to total value of 0.993 which shows high efficiency of the proposed technique. MCC is based on binary two class classification and results are true or false and contributes to maximum value (1.000) as an average and ROC(Receiver Operating Characteristics) area measurement is One of the most important values output by $WEKA$. They give you an idea of how the classifiers are performing in general and in the given case it shows average performance as 1.000 which means 100%.

VI. CONCLUSION

Machine Learning is an emerging field in the strategy decision making for Higher Educational Institutions. This study investigate the usage of machine learning algorithms to predict whether no detriment policy will be applicable to student due to COVID-19 or student performance will not be affected due to online learning by addressing all the desire measures. The study was conduct to test the performance of different machine learning algorithms which are Support

TABLE III
ALGORITHMS COMPARISON WITH QUALITY MEASUREMENT FACTORS

Methods	Decision Tree	Support Vector Machine	Naive Bayes	Random Forest
Correlation Coefficient	0.8618	0.7203	0.7173	0.9927 Kappa
Mean Absolute Error	188.4327	301.3479	298.6501	0.005
Root Mean Square Error	299.63	408.7504	410.5908	0.0357
Relative Absolute Error	34.1278 %	54.5784 %	54.0897 %	10.9071 %
Root Relative Absolute Error	50.7935 %	69.2917 %	69.6037 %	23.7473 %
Total Number of Instances				150

TABLE IV
ACCURACY BY CLASSIFICATION

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.014	0.999	1.000	1.000	0.993	1.000	1.000	No detriment Applicable
0.983	0.000	1.000	0.983	0.991	0.991	1.000	1.000	No detriment Not Applicable
0.999	0.013	0.999	0.999	0.999	0.993	1.000	1.000	Weight Average

Vector Machine, Random Forest, Decision Tree and Navie Bayes. The study show that Random Forest algorithm is able to provide high performance. The future work can include the implementing of proposed methodology for wide scale to measure performance at larger scales which can help to take strategies decision during the pandemic.

REFERENCES

[1] B. Du Boulay and R. Luckin, "Modelling human teaching tactics and strategies for tutoring systems: 14 years on," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 393–404, 2016.

[2] A. Joshi, D. Allessio, J. Magee, J. Whitehill, I. Arroyo, B. Woolf, S. Sclaroff, and M. Betke, "Affect-driven learning outcomes prediction in intelligent tutoring systems," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pp. 1–5, IEEE, 2019.

[3] B. Du Boulay and R. Luckin, "Modelling human teaching tactics and strategies for tutoring systems: 14 years on," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 393–404, 2016.

[4] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis.," *Journal of educational psychology*, vol. 106, no. 4, p. 901, 2014.

[5] V. Marian, A. Shook, and S. R. Schroeder, "Bilingual two-way immersion programs benefit academic achievement," *Bilingual research journal*, vol. 36, no. 2, pp. 167–186, 2013.

[6] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, "Intelligent tutoring systems and learning outcomes: A meta-analysis.," *Journal of educational psychology*, vol. 106, no. 4, p. 901, 2014.

[7] B. Du Boulay and R. Luckin, "Modelling human teaching tactics and strategies for tutoring systems: 14 years on," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 393–404, 2016.

[8] W. M. Al-rahmi, M. S. Othman, and M. A. Musa, "The improvement of students' academic performance by using social media through collaborative learning in malaysian higher education," *Asian Social Science*, vol. 10, no. 8, p. 210, 2014.

[9] A. Mitrovic, B. Martin, and P. Suraweera, "Intelligent tutors for all: Constraint-based modeling methodology, systems and authoring," 2007.

[10] V. Aleven, I. Roll, B. M. McLaren, and K. R. Koedinger, "Help helps, but only so much: Research on help seeking with intelligent tutoring systems," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 1, pp. 205–223, 2016.

[11] M. E. Porter *et al.*, "What is strategy?," *Harvard business review*, vol. 74, no. 6, pp. 61–78, 1996.

[12] P. T. Ward, D. J. Bickford, and G. K. Leong, "Configurations of manufacturing strategy, business strategy, environment and structure," *Journal of management*, vol. 22, no. 4, pp. 597–626, 1996.

[13] P. Kotler and P. E. Murphy, "Strategic planning for higher education," *The journal of higher education*, vol. 52, no. 5, pp. 470–489, 1981.

[14] P. F. Drucker, *Managing for the Future*. Routledge, 1993.

[15] E. Jimenez, V. Nguyen, and H. A. Patrinos, *Stuck in the middle? Human capital development and economic growth in Malaysia and Thailand*. The World Bank, 2012.

[16] Y. V. Nieto, V. García-Díaz, and C. E. Montenegro, "Decision-making model at higher educational institutions based on machine learning.," *J. UCS*, vol. 25, no. 10, pp. 1301–1322, 2019.

[17] E. López-Santana, W. Rodríguez-Vásquez, and G. Méndez-Giraldo, "A hybrid expert system, clustering and ant colony optimization approach for scheduling and routing problem in courier services," *International Journal of Industrial Engineering Computations*, vol. 9, no. 3, pp. 369–396, 2018.

[18] C. Stoean and R. Stoean, "Post-evolution of variable-length class prototypes to unlock decision making within support vector machines," *Applied Soft Computing*, vol. 25, pp. 159–173, 2014.

[19] T. Morrow, "Personalizing education with algorithmic course selection," 2017.

[20] D. Williams and M. Kennedy, "Towards a model of decision-making for systems requirements engineering process management," in *International System Dynamics 2000 Conference, Bergen, Norway*, pp. 6–10, 2000.

[21] N. Werghi and F. K. Kamoun, "A decision-tree-based system for student academic advising and planning in information systems programmes," *International Journal of Business Information Systems*, vol. 5, no. 1, pp. 1–18, 2010.

[22] C. C. White, "A survey on the integration of decision analysis and expert systems for decision support," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, pp. 358–364, 1990.

[23] M. Zorrilla, D. García, and E. Álvarez, "A decision support system to improve e-learning environments," in *Proceedings of the 2010 EDBT/ICDT Workshops*, pp. 1–8, 2010.

[24] X.-Y. Liu, "Private colleges teachers evaluation system based on support vector machine (svm)," in *2015 International conference on Applied Science and Engineering Innovation*, Atlantis Press, 2015.

[25] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, and K. L. Addison, "A machine learning framework to identify students at risk of adverse academic outcomes," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1909–1918, 2015.

[26] M. Tan and P. Shao, "Prediction of student dropout in e-learning program through the use of machine learning method.," *International Journal of Emerging Technologies in Learning*, vol. 10, no. 1, 2015.

[27] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, and K. L. Addison, "A machine learning framework to identify students at risk of adverse academic outcomes," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1909–1918, 2015.

[28] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, "Supporting academic decision making at higher educational institutions using machine learning-based algorithms," *Soft Computing*, vol. 23, no. 12, pp. 4145–4153, 2019.

[29] A. Oztekin, D. Delen, A. Turkyilmaz, and S. Zaim, "A machine learning-

based usability evaluation method for elearning systems," *Decision Support Systems*, vol. 56, pp. 63–73, 2013.

[30] A.-S. Hoffait and M. Schyns, "Early detection of university students with potential difficulties," *Decision Support Systems*, vol. 101, pp. 1–11, 2017.

[31] A. Maseleno, N. Sabani, M. Huda, R. Ahmad, K. A. Jasmi, and B. Basiron, "Demystifying learning analytics in personalised learning," *International Journal of Engineering & Technology*, vol. 7, no. 3, pp. 1124–1129, 2018.

[32] S. Suthaharan, "Machine learning models and algorithms for big data classification," *Integr. Ser. Inf. Syst*, vol. 36, pp. 1–12, 2016.

[33] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.

[34] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An ensemble random forest algorithm for insurance big data analysis," *Ieee access*, vol. 5, pp. 16568–16575, 2017.

[35] S. Amini, S. Homayouni, A. Safari, and A. A. Darvishsefat, "Object-based classification of hyperspectral data using random forest algorithm," *Geo-spatial information science*, vol. 21, no. 2, pp. 127–138, 2018.

[36] K. Jain, "Clustering algorithm of data mining to detect network statistics," *JIMS8I-International Journal of Information Communication and Computing Technology*, vol. 5, no. 1, pp. 268–274, 2017.

[37] V. Jain, A. Phophalia, and J. S. Bhatt, "Investigation of a joint splitting criteria for decision tree classifier use of information gain and gini index," in *TENCON 2018-2018 IEEE Region 10 Conference*, pp. 2187–2192, IEEE, 2018.

[38] J. Tanha, M. van Someren, and H. Afsarmanesh, "Semi-supervised self-training for decision tree classifiers," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 355–370, 2017.

[39] F. Tian, X. Cheng, G. Meng, and Y. Xu, "Research on flight phase division based on decision tree classifier," in *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*, pp. 372–375, IEEE, 2017.

[40] M. M. Saritas and A. Yasar, "Performance analysis of ann and naive bayes classification algorithm for data classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 7, no. 2, pp. 88–91, 2019.

[41] J. C. Griffis, J. B. Allendorfer, and J. P. Szaflarski, "Voxel-based gaussian naïve bayes classification of ischemic stroke lesions in individual t1-weighted mri scans," *Journal of neuroscience methods*, vol. 257, pp. 97–108, 2016.