Estimation of Learners' Levels of Adaptability in Online Education Using Imbalanced Dataset.

Manjima Sree
M.Sc AI Student
PG Department of CA & AI
Saintgits college of applied science
Kottayam, Kerala
manjima.mscai2224@saintgits.org

Jobel John James
M.Sc AI Student
PG Department of CA & AI
Saintgits college of applied science
Kottayam, Kerala
jobel.mscai2224@saintgits.org

Alphy Shaji
M.Sc AI Student
PG Department of CA & AI
Saintgits college of applied science
Kottayam,Kerala
alphy.mscai2224@saintgits.org

Dr.Ambily Merlin Kuruvilla
Associate Professor
PG Department of CA & AI
Saintgits college of applied science
Kottayam, Kerala
ambily.merin@saintgits.org

2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE) | 979-8-3503-4167-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/RASSE60029.2023.10363543

Abstract: The COVID-19 epidemic has had a huge impact on education, causing a quick move to online learning environments. Students had to adjust to a virtual learning environment as a result of this move, which brought new obstacles for them. The study looks at how adaptable students are in online learning and how it affects their academic performance and general learning process. This study aims to assess students' levels of adaptation to online learning settings and to pinpoint the factors that either support or hinder students' ability to adapt to this new learning environment. According to preliminary data, with respect to personal circumstances, students' adaptation in online learning differs greatly. The dataset used for the study is students_adaptability_level_online_education.csv is taken from Kaggle Machine Learning repository. The minority class in the dataset is equalized using the Synthetic Minority Oversampling Technique (SMOTE), as it comprises an unbalanced set of values. In order to equalize the distribution of data, the values for the minority class are increased at random. Dataset has been employed to predict the degree of student adaptability to online education through a number of machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbours (KNN) and xgb.XGBClassifier(). The Random Forest classifier had the greatest accuracy (92%), compared to those that were employed. The demand for customized and adaptable learning experiences grows as online education continues to transform the traditional learning environment. This study addresses the idea of student adaptivity in the education with an emphasis on how it may significantly enhance learning results for students from a variety of backgrounds and learning styles.

Keywords: online education, adaptability, students, COVID-19, virtual learning, challenges, Machine Learning, Prediction, Random Forest, Classification.

I. Introduction

As modernism continues to advance, progress is observed across various fields. However, the extent of our achievements in online education remains a pertinent question. “Have we acquired the necessary skills to effectively adapt to online learning?” This study aims to examine the current phase of online education and predict students' adaptability to the online mode of learning. Over the past few years, traditional education has undergone significant transformations. The Covid-19 [3,11] epidemic has drawn attention in particular to the dependence on online courses as the only way to further one's education. The advent of the internet and technological advancements has revolutionized the way education is delivered and accessed. One of the significant outcomes of this digital transformation is the emergence of online education. Online learning provides individuals with flexibility, convenience, and access to many educational possibilities that were previously unavailable to them. Particularly in recent years, the global education landscape [3,11] has witnessed a significant shift towards online learning, catalyzed further by the Covid-19 pandemic.

Being able to adjust to online learning become crucial for students to successfully navigate the current educational landscape. Adapting to online education involves not only embracing new technologies and digital tools but also adjusting one's learning style and study habits to the virtual environment. It requires a shift in mindset, self-discipline, and effective time management.

The analysis explores the potential moderating effects of factors such as gender, age, education level, location, class-time, internet connection quality, institution type (government/non-government), self-learning management system (LMS), and students' adaptability levels. By gaining insights into these factors, we can identify strategies to enhance our adaptability to online education and maximize our learning outcomes. This paper aims to explore the concept of students' adaptability to online education, examining the factors that influence their ability to thrive in this mode of learning. We may better prepare ourselves to take advantage of online education's advantages and get past any hurdles by being aware of its challenges and opportunities.

According to the study's findings, 92% of students demonstrate a good level of adaptation to online learning. However, a number of challenges prevent a seamless shift to online learning during the pandemic, including restricted access to a reliable internet connection, a lack of digital literacy, problems with electricity and network connectivity, a lack of suitable training, and resistance to the change. Prior to the Covid-19 outbreak [3,11], students had limited awareness of online education, but the sudden impact of the pandemic has left them grappling to adapt to this new learning format. Significantly, differences in student performance and satisfaction confirm the many benefits of online education. Addressing these challenges and improving the online education system becomes a crucial objective moving forward. The ultimate objective of this study is to offer insightful analysis and suggestions that will help students get the most out of their online learning. By developing the necessary skills and mindset for online education, we can take full advantage of the opportunities it presents, ensuring our academic success and personal growth in the digital age.

The focus of this project is to explore the use of data mining in implementing student adaptivity in the online education domain. By analyzing comprehensive datasets encompassing students' interactions, performance metrics, and learning behaviors, the study aim to construct adaptive models that can efficiently identify areas where students excel or struggle, and subsequently, tailor educational content and strategies accordingly.

II. Literature review

The present scenario clearly demonstrates the significance of the online education system. Technology breakthroughs have made it possible for us to create online learning platforms. Technology provides a solution to the issue of virtual or remote learning. Under the circumstances, aspects of education will be included in digitalization. To deal with these changes, students must face the challenge of transitioning to online education. In the conversation that follows, we offer a summary of the lessons discovered after reading significant articles about online education.

The researchers looked at ways to improve the online education approach in [2] and [4]. Online learning has a number of benefits for students, according to Rojan et al. [2], and has a noticeable effect on how well students do and are satisfied with their education. It displays comparable performances, similar offerings, and comparable student happiness both on and off campus. 85% of students responded that they learn more in online education, according to the exploratory results of a study by William et al. [4] that focused on formative assessment for greater learning. The main goal of the research was to enhance the teacher and student evaluation systems, as well as peer and self evaluation for both groups of people.

In their study, Rolim et al. [5] employed a Supervised ML algorithm to identify the presence of satisfactory practices, focusing on the input they gathered from the LMS courses. William et al. looked at how machine learning and data analysis in a learning management system (LMS) could improve the online education approach. [4].Researchers from [6] employed machine learning to focus on students' academic achievement. The project's goal is to create and assess the performance of a few machine learning algorithms for analysis and prediction of the student's academic achievement in the course. This paper describes the project's outcomes and evolution. According to a study by Monica et al. [7], education is going online and course materials are now available on digital platforms. They therefore conduct their analysis using Decision Trees, Support Vector Machines (SVM), Neural Networks, and Cluster Analysis. Their blend learning prediction accuracy wasn't good enough. The preliminary findings indicate that online education is superior to hybrid education. The effectiveness of machine learning in teaching was researched by Kuck et al. [8]. This study's main objective was to assess the viability of using machine learning to the field of education. The primary four types were described, Student grading comes first. By removing human biases, machine learning can rate students. They made an effort to enhance how problem-solving is evaluated in schools. Improving student retention is the second. The following step is projecting student performance, and testing students comes last.

To predict student pass rates in online education, Xiaofeng et al. [9] looked at a few specific student variables. In order to determine more crucial student characteristics influencing learning, this study attempted to forecast student pass rates and establish the most efficient machine learning method. To build a feature model, they combined three algorithms: DT, SVM, and DNN. The dropout rate is a significant issue in online education or E-learning courses, according to research by Mingjie et al. [10]. They made an effort to foresee a practical method to stop student dropouts.

The researcher has investigated that COVID-19 is a worry for international educational systems in [1] and [11]. More than 100 countries closed schools due to coronary disease. In addition to outlining a variety of obstacles that prevent students and instructors from working together in online learning environments when COVID-19 is under lockdown, their study exposes the catastrophic consequences of the corona virus on learning. Additionally, they learned that the pandemic revealed a variety of barriers to online education, such as technological barriers, individual barriers, domestic barriers, institutional barriers, communication barriers, poor electricity, network problems, a lack of adequate training, a lack of funding, resistance to change, and others. We performed study to find out whether students could adapt to online learning in this pandemic situation. Platforms for online education offer a solution to this problem. One problem, though, is how well online education does. There have so been several research on this.

III. Methodology

Three sections make up the methodology of the work, data pre-processing, balancing the imbalanced classifiers and model descriptions.

Dataset Description

A well-known Machine Learning repository for sharing and assessing datasets, Kaggle, provided the dataset used in this study to assess students' ability to adapt to online learning. Kaggle provides a wide range of datasets that have been contributed by individuals, organizations, and scholars who operate in various sectors. The dataset employed must be consistent with the goals and parameters of our investigation. Researchers thoroughly evaluated the dataset's dependability and quality. To make sure the dataset satisfies the needs of our study, this evaluation entails evaluating for completeness, correctness, and relevance.

The dataset used for the study is students_adaptability_level_online_education.csv. The dataset consists of the following fields,

1. Gender: This attribute indicates the gender of the individual participating in the online education. It is a categorical variable with options such as Girl and Boy.

2. Age: This attribute represents the age of the individual participating in online education.

3. Education Level: This characteristic displays the individual's greatest level of schooling. It is a categorical variable with choices like school, college, and university.

4. Institution Type: This attribute describes the type of educational institution the individual is associated with, such as Government or Non-Government.

5. IT Student: This attribute is a binary variable that indicates whether the individual is an Information Technology (IT) student or not. It may be represented as "Yes" or "No."

6. Location: This attribute represents the geographical location of the individual participating in online education.

7. Load-shedding: This attribute indicates whether the location experiences load-shedding (power outages) or not. It may be represented as "Low" or "High".

8. Financial Condition: This attribute reflects the financial situation of the individual, which can be classified into categories like "Poor," "Mid," and "Rich."

9. Internet Type: This attribute describes the type of internet connection the individual uses, such as Wifi or Mobile Data.

10. Network Type: This attribute indicates the type of network the individual connects to for online education, such as 4G or 3G.

11. Class Duration: This attribute represents the duration of online class session.

12. Self Lms: This attribute is a binary variable that indicates whether the individual uses a self-managed Learning Management System (LMS) or a platform-provided LMS. It may be represented as "Yes" or "No."

13. Device: This attribute describes the type of device the individual uses for online education, such as Mobile, Computer or Tab.

14. Adaptivity Level: This attribute is the target variable that indicates the adaptability level of the individual to online education. The grade, which can be given in the form of "High," "Moderate," or "Low," assesses a person's capacity to successfully adjust to and perform in an online learning environment. Each row in the dataset represents a unique individual or student, and these attributes provide valuable information for analyzing the factors that contribute to their adaptability in online education settings.

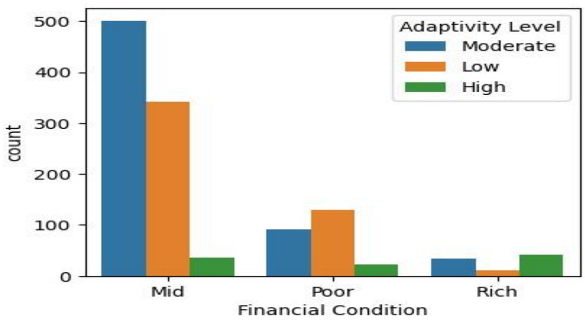

Figure1 : Students Adaptability in terms of Financial Condition (Python Generated)

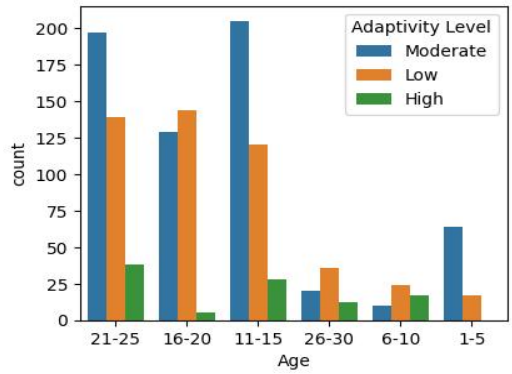

Figure2. Students Adaptability in terms of Age (Python Generated).

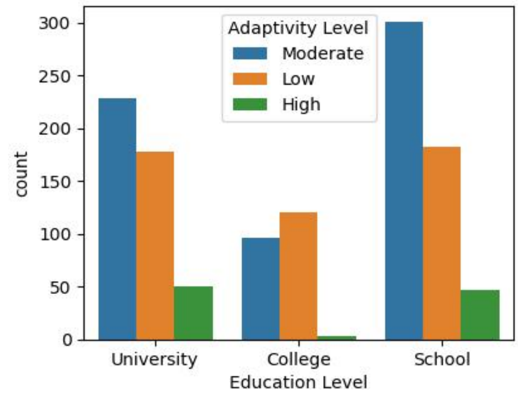

Figure3. Students Adaptability in terms of Educational Level (Python Generated).

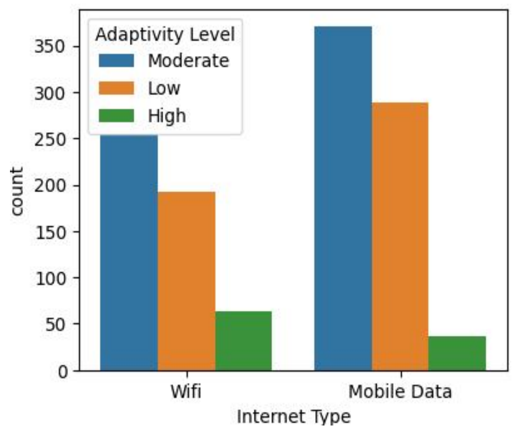

Figure 4. Students Adaptability in terms of Internet Type (Python Generated).

The dataset consists of count of various values in terms of adaptability levels of leaners. The Figure 1, Figure 2, Figure 3 and Figure 4 representing the data count in terms of High, Medium, and Low adaptability level.

A. *Data Prepossessing*

In this study before the analysis, cleaning, transforming, and organizing the data to ensure its quality and suitability for the research objective has been completed. After that, the data is normalized. Normalization rules are applied to the data set and three states are assigned to represent the low, moderate, and high adaptability level respectively. Inter Quartile Range [71,72] is employed for removing erroneous data.

B. *Balancing The Dataset using SMOTE*

Since the dataset consist of 625 Moderate values, 480 low values and 100 high values there is an imbalance in the data. To avoid the imbalance in data over sampling technique known as SMOTE [73,74] is implemented. The data values are feed to each classifier for classification after balancing the dataset.

C. *Description of Models*

There are lots of ML algorithms to predict the possible outcomes of student adaptability levels. This study have trained and tested the dataset with various ML algorithms. The algorithms that were used for prediction and analysis are K-NearestNeighbor, RandomForest, Support Vector Classifier, Logistic Regression and also XG Boost Classifier.

Since the dataset is multi labelled dataset, One Vs. Rest approach is incorporated to the binary class classification algorithms for addressing the multi class classification problem. In this approach one class is taken as positive class and all other once are together taken as negative. Each class is tested individually and the results were generated. One class is tested against all the other classes using this approach.

1) *Random Forest*

The Random Forest Classifier is a machine learning technique applied to classification issues. Multiple decision trees are combined in this ensemble learning technique to produce predictions. The random forest produces a variety of decision trees since each tree is trained using a different subset of the training data and characteristics. The incoming data is separately categorized by each tree as predictions are made, and the correct prediction is selected by majority vote. The Random Forest Classifier is renowned for its capacity to manage highly dimensional data, reduce overfitting, and produce reliable and accurate classification outcomes. Due to its efficiency and adaptability, it is frequently utilized in a variety of industries, including banking, healthcare, and picture categorization.

$$\text{norm}fi_i = \frac{fi_i}{\sum j \in features fi_i} \quad \text{Eq(1)}$$

2) *K-Nearest Neighbor*

K-Nearest Neighbor (KNN): KNN is one of the simplest ML algorithms out of the many available. The ease of quick interpretation and short calculation times of the KNN can be credited for its popularity [16]. KNN chooses k neighbors, calculates k neighbors distance functions, and then uses the distance functions to assign the class shared most frequently by its k closest neighbors. The formula used to calculate the distance vector is:

$$d(x, y)=\sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

3) *Support Vector Classifier*

A supervised learning system called a support vector classifier is employed for classification and regression issues. Many people strongly favor the support vector machine since it generates significant correctness while using less processing power. It is primarily applied to classification issues. Unsupervised, supervised, and reinforced learning are the three categories of learning. A support vector machine is correctly referred to as a selective classifier because it partitions the hyperplane. The technique generates the ideal hyperplane that, given labelled training data, accurately classifies new samples. This hyperplane is a line that, in two-dimensional space, divides a plane into two areas, one region for each class on either side. Finding an N-dimensional hyperplane that classifies the data points separately is the objective of the support vector machine method.

1. w * x - b = 1: This equation represents the decision boundary for the positive class in an SVM classifier. The vector w represents the weights (coefficients) assigned to each feature in the dataset, and x represents the feature vector of a data point. Bias is expressed by the word b. In essence, the equation says that for any data point x falling under the positive class, the value of w * x (the dot product of the weight vector and the feature vector) minus b should be greater than or equal to 1 for proper classification.

2. w * x - b = 0: The decision boundary for the difference between the positive and negative classes is represented by this equation. By doing this, it makes sure that the separation between the decision boundaries for the positive class and the negative class is maximized. Data points near this boundary are called support vectors because they have a significant influence on the optimum decision boundary selection.

3. w * x - b = -1: This equation represents the decision boundary for the negative class in an SVM classifier. It is similar to the first equation, but for the negative class, and it ensures that data points belonging to the negative class have a value of w * x minus b less than or equal to -1 for proper classification.

4. 2 / ||w||: The margin in SVM is the angle formed between the decision boundary and the support vectors. This margin will be increased as much as possible via SVM. The magnitude of the weight vector (w) has an inverse relationship with the margin. Therefore, this equation represents half of the margin, which is equal to 2 divided by the weight vector's Euclidean norm (||w||).

5. b / ||w||: The decision boundary's offset from the origin is calculated using this equation. The bias term has a value of b, and the weight vector has a magnitude of ||w||. The act of changing the decision boundary to correctly categories the data points is referred to in this statement.

These equations are essential in understanding how SVM works and how it finds the optimal hyperplane to separate different classes in the data space with the largest margin. Due to its efficiency and adaptability, SVM is frequently utilized in a range of machine learning applications.

4) *XGB Classifier*

XG Boost is a distributed gradient boosting toolkit that has been tuned for quick and scalable machine learning model training. A number of weak models' predictions are combined using this ensemble learning technique to get a stronger prediction. Extreme Gradient Boosting, or XG Boost, is one of the most well-known and widely used machine learning algorithms because it can handle large datasets and perform at the cutting edge in many machine learning tasks like classification and regression. Its effective handling of the missing values, which enables to handle real-world data with missing values without requiring a lot of pre-processing, is one of the key characteristics of XGBoost. Additionally, XGBoost includes integrated parallel processing capability, allowing you to train models on huge datasets quickly.Applications for XGBoost include click-through rate prediction, recommendation systems, and Kaggle competitions among others. Additionally, it is quite adaptable and enables performance optimization by allowing for fine-tuning of numerous model parameters.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, y_i^{(t-1)} + f_t(X_i)) + \Omega(f_t)$$

IV. RESULT AND DISCUSSION

The effectiveness of ML models can be evaluated using a variety of metrics. Precision, Recall, F1 score, and Accuracy are the most important characteristics used to assess a model's performance. The value of the confusion matrix which is generated during the testing of the model is considered to calculate the score of the precision, recall, F1-Score, and accuracy. The formulas [19] that are used in these computations are given in equations below:

$$\text{Precision} = \frac{TPV}{TPV+FPV}$$

$$\text{Recall} = \frac{TPV}{TPV+FNV}$$

$$\text{Accuracy} = \frac{TPV+TNV}{TPV+FPV+FNV+TNV}$$

$$\text{F1-Score} = \frac{2*Precision*Recall}{Precision+recall}$$

Here, True positives values are represented by TPV, true negatives values by TNV, false positives values by FPV, and false negatives values by FNV.

Table 1 shows the values for the various machine learning models used in this study in terms of accuracy, precision, recall, and F1 Score. The following algorithms were utilized in the current study: K-Nearest Neighbors, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (), and XGB Classifier. For Random Forest, accuracy is 92%. The results of this study could aid decision-makers in the education sector in better understanding the current online education system and the levels of High, Medium, and Low learners' adaptability to online learning.

Model	Class Name	Accuracy	Precision	Recall	F1 Score
RF	Low adaptability	92	0.93	0.93	0.93
	Moderate adaptability		0.93	0.85	0.89
	High adaptability		0.92	1.00	0.96
SVC	Low adaptability	80	0.85	0.77	0.81
	Moderate adaptability		0.72	0.69	0.71
	High adaptability		0.82	0.93	0.87
KNN	Low adaptability	80	0.86	0.89	0.87
	Moderate adaptability		0.89	0.72	0.80
	High adaptability		0.87	1.00	0.93
XGB	Low adaptability	92%	0.93	0.93	0.93
	Moderate adaptability		0.93	0.84	0.88
	High adaptability		0.91	0.99	0.95

TABLE1. Results of Machine Learning Models

FUTURE SCOPE

Hybridized models of Random Forest along with the nature loving algorithms like Grasshopper Optimization, Gray wolf optimizer will give better performance. The study can be extended to deep learning models like Auto Encoder, SpinalNet etc.

REFERENCES

[1] E. M. Onyema, N. C. Eucheria, F. A. Obafemi, S. Sen, F. G. Atonye, A. Sharma, and A. O. Alsayed, "Impact of coronavirus pandemic on education," Journal of Education and Practice, vol. 11, no. 13, pp. 108– 121, 2020.

[2] R. Afrouz and B. R. Crisp, "Online education in social work, effectiveness, benefits, and challenges: A scoping review," Australian Social Work, vol. 74, no. 1, pp. 55–67, 2021.

[3] M. Tabassum, S. E. Mannan, M. I. Parvej, and F. Ahmed, "Online education during covid-19 in bangladesh: University teachers' perspective," Aquademia, vol. 5, no. 1, p. ep21005, 2021.

[4] D. Wiliam, "Assessment in education: Principles, policy & practice," Assessment in Education: Principles, Policy and Practice, vol. 15, no. 3, pp. 253–257, 2008.

[5] A. P. Cavalcanti, R. F. L. de Mello, V. Rolim, M. Andre, F. Freitas, ´ and D. Gasevic, "An analysis of the use of good feedback practices in ˇ online learning courses," in 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), vol. 2161. IEEE, 2019, pp. 153–157.

[6] V. L. Uskov, J. P. Bakken, A. Byerly, and A. Shah, "Machine learningbased predictive analytics of student academic performance in stem education," in 2019 IEEE Global Engineering Education Conference (EDUCON). IEEE, 2019, pp. 1370–1376.

[7] M. Ciolacu, A. F. Tehrani, R. Beer, and H. Popp, "Education 4.0—fostering student's performance with machine learning methods," in 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, 2017, pp. 438–443.

[8] D. Kucak, V. Juri ˇ ciˇ c, and G. ambi ´ c, "Machine learning in education-a ´ survey of current research trends." Annals of DAAAM & Proceedings, vol. 29, 2018.

[9] X. Ma, Y. Yang, and Z. Zhou, "Using machine learning algorithm to predict student pass rates in online education," in Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing, 2018, pp. 156–161.

[10] 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.

[11] R. E. Baticulon, J. J. Sy, N. R. I. Alberto, M. B. C. Baron, R. E. C. Mabulay, L. G. T. Rizada, C. J. S. Tiu, C. A. Clarion, and J. C. B. Reyes, "Barriers to online learning in the time of covid-19: A national survey of medical students in the philippines," Medical science educator, pp. 1–12, 2021.

[12] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." Psychological methods, vol. 14, no. 4, p. 323, 2009.

[13] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M. Rahman, "Predicting students' performance of the private universities of bangladesh using machine learning approaches," International Journal of Advanced Computer Science and Applications, vol. 11, no. 3, pp. 672–679, 2020.

[14] "Machine learning 101: Decision tree algorithm for classification," Online: https://www.analyticsvidhya.com/blog/2021/02/machinelearning-101-decision-tree-algorithm-for-classification/, Access Date: 10/05/2021.

[15] A. A. Biswas, A. Majumder, M. J. Mia, I. Nowrin, and N. A. Ritu, "Predicting the enrollment and dropout of students in the post-graduation degree using machine learning classifier," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 11, pp. 3083–3088, 2019.

[16] "Understanding math behind knn (with codes in python)," Available Online: http://www.datascribble.com/blog/machinelearning/understanding-math-behind-knn-codes-python/, Access Date: 22/05/2021.

[17] W. S. Alaloul and A. H. Qureshi, "Data processing using artificial neural networks," in Dynamic Data Assimilation-Beating the Uncertainties. IntechOpen, 2020.

[18] J. Zhan and B. Dahal, "Using deep learning for short text understanding," Journal of Big Data, vol. 4, no. 1, pp. 1–15, 2017.

[19] A. A. Biswas, A. Majumder, M. J. Mia, R. Basri, and M. S. Zulfiker, "Career prediction with analysis of influential factors using data mining in the context of bangladesh," in Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Springer, 2021, pp. 441–451.

[20] M. J. Mia, A. A. Biswas, A. Sattar, and M. Habib, "Registration status prediction of students using machine learning in the context of private university of bangladesh," International Journal of Innovative

[21] S.-T. Wu, "Knowledge discovery using pattern taxonomy model in textmining," 2007

[22] J. Mostow and J. Beck, "Some useful tactics to modify, map and minedata from intelligent tutors," Natural Language Engineering, vol. 12,no. 02, pp. 195–208, 2006.

[23] S. K. Mohamad and Z. Tasir, "Educational data mining: A review,"Procedia-Social and Behavioral Sciences, vol. 97, pp. 320–324, 2013.

[24] R. Baker et al., "Data mining for education," International encyclopediaof education, vol. 7, pp. 112–118, 2010.

[25] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery withgenetic programming for providing feedback to courseware authors,"User Modeling and User-Adapted Interaction, vol. 14, no. 5, pp. 425–464, 2004

[26] N. S. Raghavan, "Data mining in e-commerce: A survey," Sadhana,vol. 30, no. 2-3, pp. 275–289, 2005.

[27] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, Handbook ofeducational data mining. CRC Press, 2010.

[28] R. S. Baker and K. Yacef, "The state of educational data mining in2009: A review and future visions," JEDM-Journal of Educational DataMining, vol. 1, no. 1, pp. 3–17, 2009.

[29] F. Castro, A. Vellido, `A. Nebot, and F. Mugica, "Applying data mining techniques to e-learning problems," in Evolution of teaching andlearning paradigms in intelligent environment, pp. 183–221, Springer,2007.

[30] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The kdd process forextracting useful knowledge from volumes of data," Communications ofthe ACM, vol. 39, no. 11, pp. 27–34, 1996.

[31] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, in Proc. Educational Data Mining 2009:2nd International Conf, 2009.

[32] S. L. Tanimoto, "Improving the prospects for educational data mining,"in Track on Educational Data Mining, at the Workshop on Data Miningfor User Modeling, at the 11th International Conference on UserModeling, pp. 1–6, 2007.

[33] J. E. Beck and J. Mostow, "How who should practice: Using learningdecomposition to evaluate the efficacy of different types of practice fordifferent types of students," in Intelligent tutoring systems, pp. 353–362,Springer, 2008.

[34] M. Cocea, A. Hershkovitz, and R. S. Baker, "The impact of off-task andgaming behaviors on learning: immediate or aggregate?," 2009.

[35] H. Jeong and G. Biswas, "Mining student behavior models in learning-by-teaching environments.," in EDM, pp. 127–136, Citeseer, 2008.

[36] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies,"Artificial Intelligence Review, vol. 22, no. 2, pp. 85–126, 2004.

[37] C. C. Chan, "A framework for assessing usage of web-based e-learningsystems," in Innovative Computing, Information and Control, 2007.ICICIC '07. Second International Conference on, pp. 147–147, Sept2007.

[38] M. Muehlenbrock, "Automatic action analysis in an interactive learningenvironment," in Proceedings of the 12 th International Conference onArtificial Intelligence in Education, pp. 73–80, 2005.

[39] M. Ueno, "Data mining and text mining technologies for collaborativelearning in an ilms" ssamurai"," in Advanced Learning Technologies,2004. Proceedings. IEEE International Conference on, pp. 1052–1053,IEEE, 2004.

[40] L. P. Dringus and T. Ellis, "Using data mining as a strategy for assessingasynchronous discussion forums," Computers & Education, vol. 45,no. 1, pp. 141–160, 2005.

[41] J. Chen, Q. Li, L. Wang, and W. Jia, "Automatically generating an e-textbook on the web," in Advances in Web-Based Learning–ICWL 2004,pp. 35–42, Springer, 2004.

[42] J. Tane, C. Schmitz, and G. Stumme, "Semantic resource managementfor the web: an e-learning application," in Proceedings of the 13thinternational World Wide Web conference on Alternate track papers &posters, pp. 1–10, ACM, 2004.

[43] C. Tang, R. W. Lau, Q. Li, H. Yin, T. Li, and D. Kilis, "Personalizedcourseware construction based on web data mining," in Web

Informa-tion Systems Engineering, 2000. Proceedings of the First InternationalConference on, vol. 2, pp. 204–211, IEEE, 2000.

[44] J. Scott, Social network analysis. Sage, 2012.

[45] P. Reyes and P. Tchounikine, "Mining learning groups' activities inforum-type tools," in Proceedings of th 2005 conference on Computersupport for collaborative learning: learning 2005: the next 10 years!,pp. 509–513, International Society of the Learning Sciences, 2005.

[46] C. Romero, S. Ventura, P. G. Espejo, and C. Herv´as, "Data miningalgorithms to classify students.," in EDM, pp. 8–17, 2008.

[47] A. P. Bradley, "The use of the area under the roc curve in the evaluationof machine learning algorithms," Pattern recognition, vol. 30, no. 7,pp. 1145–1159, 1997.

[48] B. Liu, Web data mining: exploring hyperlinks, contents, and usage data.Springer Science & Business Media, 2007.

[49] C. R. Beal, L. Qu, and H. Lee, "Classifying learner engagement throughintegration of multiple data sources," in Proceedings of the NationalConference on Artificial Intelligence, vol. 21, p. 151, Menlo Park, CA;Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[50] S. Amershi and C. Conati, "Automatic recognition of learner groupsin exploratory learning environments," in Intelligent Tutoring Systems,pp. 463–472, Springer, 2006.

[51] A. Merceron and K. Yacef, "Interestingness measures for associationsrules in educational data.," EDM, vol. 8, pp. 57–66, 2008.

[52] C. Wallace, K. B. Korb, and H. Dai, "Causal discovery via mml," inICML, vol. 96, pp. 516–524, Citeseer, 1996.

[53] A. Hershkovitz and R. Nachmias, "Developing a log-based motivationmeasuring tool.," in EDM, pp. 226–233, Citeseer, 2008.

[54] J. Kay, N. Maisonneuve, K. Yacef, and P. Reimann, "The big five andvisualisations of team work activity," in Intelligent tutoring systems,pp. 197–206, Springer, 2006.

[55] R. S. d Baker, A. T. Corbett, and V. Aleven, "More accurate studentmodeling through contextual estimation of slip and guess probabilitiesin bayesian knowledge tracing," in Intelligent Tutoring Systems, pp. 406–415, Springer, 2008.

[56] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling theacquisition of procedural knowledge," User modeling and user-adaptedinteraction, vol. 4, no. 4, pp. 253–278, 1994.

[57] R. Baker et al., "Data mining for education," International encyclopediaof education, vol. 7, pp. 112–118, 2010.

[58] C. Romero, S. Ventura, and P. De Bra, "Knowledge discovery withgenetic programming for providing feedback to courseware authors,"User Modeling and User-Adapted Interaction, vol. 14, no. 5, pp. 425–464, 2004.

[59] N. S. Raghavan, "Data mining in e-commerce: A survey," Sadhana,vol. 30, no. 2-3, pp. 275–289, 2005.

[60] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. Baker, Handbook ofeducational data mining. CRC Press, 2010.

[61] T. Denley, Austin Peay State University: Degree Compass, 2012, [online] Available: http://www.educause.edu/ero/article/austin-peay-state-university-degree-compass Google Scholar.

[62] M. F. M. Mohsin, N. M. Norwawi, C. F. Hibadullah and M. H. A. Wahab, "Mining the student programming performance using rough set", Int. Conf. Intell. Syst. Knowl. Eng. (ISKE), pp. 478-483, Nov. 2010.

[63] O. R. Zaıane, "Building a recommender agent for e-learning systems", Proc. Int. Conf. Comput. Edu., pp. 55-59, Dec. 2002.

[64] R. S. J. D. Baker, A. T. Corbett and A. Z. Wagner, "Off-task behavior in the cognitive tutor classroom: When students 'game the system", Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 383-390, Apr. 2004.

[65] P. Brusilovsky and C. Peylo, "Adaptive and intelligent Web-based educational systems", Int. J. Artif. Intell. Edu., vol. 13, no. 2, pp. 159-172, 2003.

[66] N. A. Rashid, M. N. Taib, S. Lias, N. Sulaiman, Z. H. Murat and R. S. S. A. Kadir, "Learners' learning style classification related to IQ and stress based on EEG", Proc.-Social Behavioral Sci., vol. 29, pp. 1061-1070, 2011.

[67] Q. Zheng, J. Ding, J. Du and F. Tian, "Assessing method for E-learner clustering", 11th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD), pp. 979-983, Apr. 2007.

[68] P. Dráždilová, J. Martinovic, K. Slaninová and V. Snášel, "Analysis of relations in eLearning", Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol. (WI-IAT), pp. 373-376, 2008.

[69] F. Tian, S. Wang, C. Zheng and Q. Zheng, "Research on e-learner personality grouping based on fuzzy clustering analysis", Proc. 12th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD), pp. 1035-1040, Apr. 2008.

[70] J. Chen, K. Huang, F. Wang and H. Wang, "E-learning behavior analysis based on fuzzy clustering", Proc. 3rd Int. Conf. Genet. Evol. Comput. (WGEC), pp. 863-866, Oct. 2009.

[71] Ambily Merlin Kuruvilla, Dr. N.V. Balaji, "Improved Artificial Neural Network Through Metaheuristic Methods And Rough Set Theory For Modern Medical Diagnosis", Indian Journal of Computer Science and Engineering (IJCSE) DOI: 10.21817/indjcse/2021/v12i4/211204161 Vol. 12 No. 4 Jul-Aug 2021 946.

[72] Ambily Merlin Kuruvilla, Dr. N.V. Balaji, "Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach.", IOP Conference Series: Materials Science and Engineering, DOI 10.1088/1757-899X/1085/1/012028

[73] Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, P. (2000). SMOTE: Synthetic Minority Over-sampling TEchnique. In International Conference of Knowledge Based Computer Systems, pp. 46–57. National Center for Software Technology, Mumbai, India, Allied Press.

[74] Fernandez, A. *et al.* SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* 61, 863–905 (2018).