International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.41, September 2023 – www.ijais.org

Machine Learning-based E-Learners' Engagement Level Prediction using Benchmark Datasets

God'swill Theophilus
Department of Computer Science, Faculty of Computing, Federal University of Lafia
P.M.B 146, Nigeria

Christopher Ifeanyi Eke
Department of Computer Science, Faculty of Computing, Federal University of Lafia
P.M.B 146, Nigeria

ABSTRACT
The wide adoption of e-learning especially during and after the pandemic has given rise to the concern of learners' motivation and involvement. E-leaner engagement level recognition over time has become critical since there is little to no physical interaction. In this paper, a benchmark dataset was utilized in predicting learners' engagement levels in a blended e-learning system. Information Gain feature ranker was leveraged to ascertain the significance of the features. This study performed a comparative study on some machine learning algorithms including; Decision Tree, Naïve Bayes, Random Forest, Logistics Regression, Stochastic Gradient Descent, LogitBoost, Sequential Minimal Optimization, Voted Perceptron, and AdaptiveBoost. Each model was accessed using the 10-fold cross-validation. We measure the performance of the models before and after feature selection. The predictive results show that Sequential Minimal Optimization outperformed other models by attaining an accuracy of 90% with precision, recall, and f-measure values of 0.895, 0.897, and 0.895 respectively.

Keywords
Data science applications in education; E-learning; Machine learning; Engagement prediction; Learning strategies.

1. INTRODUCTION
The rapid spread of the threatening Coronavirus is causing the world to experience an unprecedented situation, many sectors including the educational sector are characterized by uncertainty about the future. During the worldwide lockdown, the academic sector had to immediately adapt to this new situation and continue functioning [1]. The pronouncement of COVID-19 as a global plague by the World Health Organization (WHO) made it essential for many countries to apply safety efforts directed towards containing the virus-related epidemic, coercing most if not all organizations to close up and have their operations carried out from distant, institutions of learning were not left out in this shift in operation [2]. Consequently, the majority of scholastic institutes were left with no option but to continue functioning despite the situation being faced globally. The switch from the traditional system of learning to online systems which usually involves careful planning and sound training for a good period of time is now taking place all of a sudden. With such an unprepared transition, one is sure that in the nearest future questions with regards to conveyance of standard education will be raised by many [3].

Several reasons exist for the adoption of e-learning during the coronavirus pandemic, few among these are efficiency and high speed of delivery, conductivity and rationality, advanced knowledge absorbing capabilities, reduced overhead cost and time, an incredible opportunity for investment, adaptability, and ability to reach learners irrespective of location. Khandelwal and Kumar [4] assert that a noteworthy advantage of e-learning is the fact that it is limited neither by time nor by geography. Along with such an unprecedented change in learning during the pandemic which resulted in the total involvement of distance learning machineries comes a higher need for improving such pedagogical systems and with the wide adoption of e-learning comes the concern for the performance of e-learners as the quality of graduates produced through this system raises the question of competence and excellence. Some e-learning systems have little or no supervision, as such one cannot be confident of the results recorded by these systems. Learners' motivation and engagement also become a concern. With the absence of face-to-face interaction, the motivation and engagement level of these learners tend to dwindle which lays a foundation for low performance. Educational Data Mining (EDM) and Machine Learning (ML) are being employed by researchers to improve the performance and functionality of e-learning systems.

Educational Data Mining (EDM) is a noticeable sphere of research. The goals of EDM comprise appraisal of knowledge and educational strategies, devising a fit atmosphere for learners based on their learning style, supplying feedback to learners and instructors as well as uncovering odd learning patterns in such systems [5]. Machine learning can simply be defined as employing data and statistics into composing significant likelihoods about an occurrence. This is akin to learning in humans. Machine learning algorithms search for meaningful relationships within a body of data and try to match inputs with outputs [6].

One of the widely used applications of machine learning in e-learning is classification. Several aspects of classification have been employed in enhancing e-learning systems, one such aspect is classifying learners based on their engagement level. Learner's engagement level provides insight into their academic performance and several unique methodologies have gained relevance in this area [7]. Rastrollo-Guerrero [8] asserted that forecasting students' performance is a very relevant aspect within the context of learning environments such as schools and universities, it provides a firm foundation upon which efficient structures that promotes learning are proposed and reduces the dropout rate. Numerous researches correlating to the prediction of students' engagement have over time been carried out entirely for the advantage of educational instructors and administrators. Leelavathy et al., [9] discovered that in a dynamic learning environment, students tend to suffer the loss of inspiration easily, an occurrence higher in systems that are not designed to meet the peculiar needs of learners. In a conventional system of learning for instance, instructors adopt several strategies such as serving tests and examinations, marking attendance, and grading notes in order to assess learners' degree of interest, engagement, and depth of understanding [10], whereas in e-learning there is little or no physical contact with learners which makes it difficult to assess students, as such data provides a good basis for measuring performance and engagement. The undeniable need to spot students with the possibility of performing below normal is very important, especially in e-learning, this is necessary to avoid expulsion or dropout in the process of learning [11]. The following are the contributions of this study to the body of knowledge;

1. Literature investigation: This research delivers an all-inclusive investigation of literatures in the field of e-learners' engagement level prediction with focus on data collection, data pre-processing, feature extraction and selection, model development and evaluation techniques
2. Suggests suitable features for engagement level prediction: The study suggest features that are fit for the development of an efficient model for the purpose of engagement level prediction in a blended e-learning system.
3. Proposes an engagement level prediction model: the study proposes a model that is more efficient when compared to baseline models for the prediction of leaners engagement. The proposed model is less complex and more accurate.
4. Using the sequential minimal optimization classifier, the engagement level prediction model achieved the highest precision of 0.897, recall 0.897 and f-measure 0.897. The evaluation shows the significance of the suggested framework for leaners' engagement level prediction after comparing the acquired findings for the proposed three baseline techniques.

The rest of this paper is structured thus: section 2 provides a detailed review of recent literatures in the field of e-learners' engagement level prediction, section 3 outlines the methodology employed in this study and further provides a brief description of the machine learning algorithms employed in this investigation. While section 4 provides the experimental setup for the research process, section 5 supplies the result of the experiment and section 6 concludes the researches and provides insight into future research interest.

2. REVIEW OF RELATED WORKS

The ability of a computer to carry out assignments after finding out some guidelines from some predefined sets of data is referred to as machine learning. We describe machine learning prediction as the prophetic dimension of computer science. Prophetic computer science provides insight into future events before they occur providing a good base for recommendation and counselling. The ability of computers to prophesy has contributed greatly to the overall advancement in technology, reducing risk and maximizing performance. There are several models which are employed in the prophetic schooling of computers, however, a broad classification of these models will give rise to two classes namely: shallow learning models and deep learning models.

Gorgun et al., [12] researched 'Predicting Cognitive Engagement in Online Course Discussion Forums' and tested the predictive performance of 3 shallow learning models: Decision Tree, Random Forest, and Support Vector Machine. The dataset was collected over an online discussion forum which was utilized to provide classes at a prestigious Canadian college of education. The dataset contained 4,217 postings that 111 students had written. Coh-Metrix 3.0 was leveraged to pre-process the data. The research was carried out in Python (Version 3.8.8) using the sklearn and mlxtend packages and the 10-fold cross-validation was employed in schooling the models. Precision, accuracy, recall, and f-1 measures were used to ascertain the performance of the models. Decision Tree provided an accuracy of 71%, Random Forest supplied an accuracy of 73% while Support Vector Machine outperform the others with an accuracy of 83%.

Similarly, Jawad et al., [13] trained Random Forest to predict learners' engagement levels in an online learning system. The dataset employed for the study was a collection of information from 22 Open University modules and it includes 32,593 students' demographics as well as aggregated information on their assessment results and clickstreams—a record of how they interacted with the university's virtual learning environment. The daily summary of the clickstream data has a total of 10,655,280 items. Because the dataset was imbalanced, SMOTE data balancing technique was employed to enhance the predictive performance of the models over the data. The data were divided into training and testing sets, with 667% being employed for training and 33% for testing. Python together with the sklearn library was employed for model development. To measure the performance of the trained model, accuracy, ROC-AUC, precision-recall AUC, and f-1 score were employed for evaluation. Random Forest provided the best accuracy level of 84%.

Hussain et al., [10] tested the ability of the Decision Tree, J48, Classification and Regression Tree (CART), JRIP Decision Rules, Gradient Boosting Trees (GBT), and Naïve Bayes Classifiers (NBC) in predicting student engagement in an e-learning system. The study employed a set of data extracted from the module of social science courses taken by 384 students. 3 features were employed for the study; demographic (highest education level), performance (final results and score on the assessment), and learning behaviour (number of clicks on VLE activities). The data was further pre-processed using MATLAB to change the format of the data to one that can be employed for the predictive process. High Engagement and Low Engagement were used as labels in classifying the learners. Decision Tree provided the best accuracy of 85.91% followed by Naïve Bayes which provided an accuracy of 82.93%.

The ability of LightGBM, Random Forest, CatBoost, XGBoost, and Multi-Layer Perceptron to predict learners' engagement was tested by Alruwais and Zakariah [14]. Some datasets were gathered from Open University in British which comprises seven selected modules and learners' engagement with the Virtual Learning Environment (VLE). The dataset contained seven features; studentInfo, studentAssessment, assessment, studentVLE, studentRegistration, VLE, and courses. The dataset also contained details of students' location, age, disability, education level, and gender. The dataset was pre-processed to remove missing values and normalized. The researchers claimed that after pre-processing they were left with 34 features. Accuracy, precision, recall, and f-measure were employed to ascertain the predictive performance of the trained models. LightGBM outperformed the other trained models with an accuracy of 92%, a precision of 0.94, and a recall of 0.93.

Similarly, with the aim of developing a model that predicts students' engagement level in a blended e-learning system, Buraimoh [5] conducted extensive research using Logistic Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forests (RF), Gradient Boosting Trees (GBT), Multilayer Perceptron (MP), and Linear Discriminant Analysis (LDA). They accessed some data made available on Kaggle. The data was originally sourced from Kalboard 360 LMS with the use of xAPI. The data consisted of 500 samples and 16 instances. Precision, recall, f-measure, and Area Under Curve (AUC) were used for evaluating the performance of the models. Random Forest provided the highest accuracy with an accuracy value of 90%. Decision Tree, Random Forests and Multilayer Perceptron have higher precision scores with precision values of 0.89 each. Random Forests and Multilayer Perceptron had the highest recall of 0.89 followed by Decision Tree with a recall of 0.88.

This research is different from the existing related work in the study domain because it focuses on predicting learners' level of engagement employing benchmark datasets. In the context of machine learning (ML), benchmarking refers to the process of assessing and comparing the performance of different ML methods in terms of their ability to learn patterns from standardized datasets known as benchmarks [15]. Benchmarking serves as a way to verify that a new method functions as intended by confirming its capability to reliably identify simple patterns already recognized by existing methods [16]. Another perspective on benchmarking involves a more rigorous evaluation, aiming to identify the strengths and weaknesses of a particular methodology in comparison to others. This assessment can encompass various evaluation metrics such as signal detection capability, prediction accuracy, computational complexity, and model interpretability. Employing benchmarking in this manner is crucial for demonstrating novel methodological advancements or aiding in the selection of an appropriate ML method for a specific problem.

3. METHODOLOGY

This section provides an overview of the research process employed in carrying out this research. We employ an experimental quantitative research approach for this study. A quantitative study aims to address a problem by using numerical data. It involves collecting, analysing, and experimenting with data to draw a conclusion. The steps employed in carrying out this research is outlined in the figure below.

3.1 Data Acquisition

This research employs a dataset gathered from the University of Western Ontario (Canada) Learning Management System (LMS) which delivers its contents in a blended learning form. The data was made available and accessible by Moubayed et al., [17] after employing it for a clustering analysis.

The dataset is made up of 486 instances and has 12 features. The features are grouped into effort and interaction. The effort category is made up of the assignment 1 lateness indicator, assignment 2 lateness indicator, assignment 3 lateness indicator, assignment 1 duration to submit, assignment 2 duration to submit, assignment 3 duration to submit, and average assignment duration to submit while the interaction category consists of several logins, number of content reads, number of forums reads, number of forum posts, and number of quiz reviews. The dataset however neither provided demographical information nor the courses or topics taught. Moubayed et al., [17] opined that this was so due to the privacy laws the university was observing.

3.2 Data Pre-processing

Data processing is a significant aspect of model development. Data acquired in their raw form contain noise and anomalies which can affect the performance of the model being schooled [18]. Missing values and white spaces are some common irregularities with raw data. Some techniques the researchers employed in cleaning the data include:

Data normalization: This is a pre-processing technique applied to numerical features before applying classification or clustering algorithms that are mainly designed to handle numerical features. The reason behind the importance of the normalization process is to avoid a number of the considered features concealing the effect of others, particularly when features have different varying ranges

Data discretization: Data discretization is used in pre-processing numeric values. Not every learning method can handle numeric values. Sometimes the learning methods may not produce the exact output when dealing with such values. This technique is applied sometimes to meet the requirement of input of models, such as Naive Bayes, which require its input to be countable.

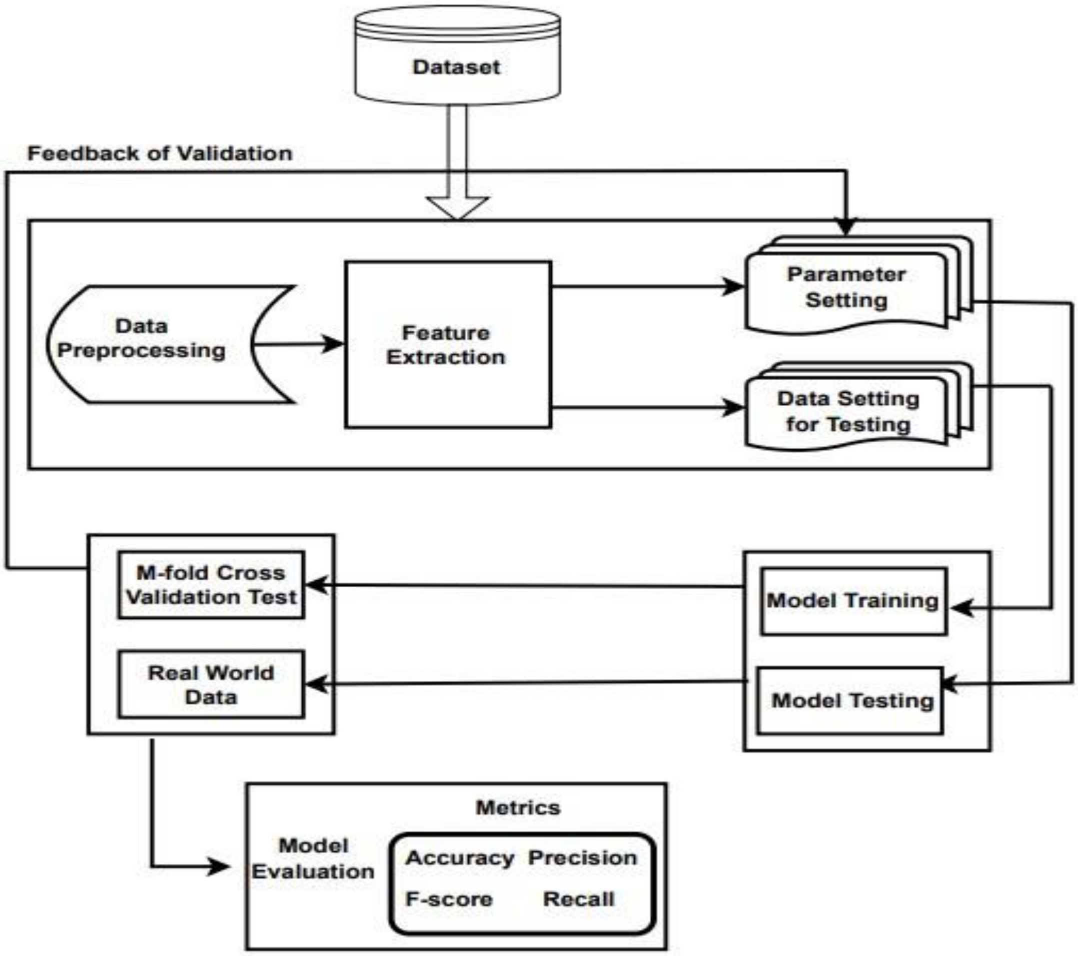

Figure 1. Engagement level prediction framework

Missing value: Missing value is a datum that has not been stored or gathered due to a faulty sampling process, cost restrictions or limitations in the acquisition process. Missing values cannot be avoided in data analysis, and they tend to create severe difficulties for practitioners. Missing value are generated due to several reasons such as human errors, equipment faults, data unavailability and data not being up-to-date or inconsistent with other existing data.

3.3 Feature Extraction

A critical step in preparing and transforming raw data into a format appropriate for machine learning algorithms is feature extraction [19]. In order to represent and capture the information that is essential for model training and analysis, a selected group of pertinent features (variables or attributes) from the original data must be chosen or created. By decreasing the amount of dimension of the data while keeping its valuable characteristics, feature extraction helps to improve the effectiveness, accuracy, and interpretability of machine learning models [20]. This research employs Information Gain to evaluate the significance of the features, which allows the researchers to identify the features that are most important to include in the schooling process of the prediction models in order to reduce complexity and enhance data quality for better results.

Information Gain feature assessment ranks features by calculating the value of individual feature by computing the entropy in the relation. This provides insights into the significance or insignificance of a feature as it relates to building a model especially when efficiency and effectiveness are necessary.

$$IG = \mathrm{H}\,(\mathrm{D},\mathrm{F}) - \sum\nolimits_{v \in values(F)} \frac{Dv}{D} \,.\, H(Dv) \qquad 3.0$$

Where H(D,F) is the Information Gain of feature F in dataset D. H(D) is the entropy of the original dataset D. Values (F) represents the possible values or categories of features F. D_v is the subset of dataset D where feature F takes V and D is the total number of instances in dataset D.

After applying the Information Gain feature ranker we find following features important in predicting learners' engagement; number of logins, number of contents reads, number of forum reads, number of quiz reviews, assignment 1 lateness indicator, assignment 2 lateness indicator, assignment 3 lateness indicator, assignment 1 duration to submit, assignment 2 duration to submit, assignment 3 duration to submit and average assignment duration to submit.

3.4 Construction of the Classification Models

This section focuses on the machine learning-based categorization algorithm used in this study. Following feature extraction, the model training stage of ML incorporates an algorithm that learns from data and produces predictions. A machine learning algorithm created for the prediction of leaners engagement was built using the dataset's attributes. This algorithm is specifically tasked with classifying leaners base on their interaction and involvement with the learning system. To find the best classifier for engagement level prediction, a number of classifiers, including DT, RF, AB, LB, VP, SGD, NB and SMO, were tested through a variety of tests.

The process of choosing the best classifier for a given dataset, however, is very difficult. Given the diversity of learning strategies used across various application areas, prior research [20] often tests two or more machine learning algorithms to find the top performer. Therefore, nine different classifiers—Decision tree, LogitBoast, Adaptive boast, Random forest, Voted perceptron, Logistic regression, Stochastic gradient descent Sequencial minimal optimization, and Random Forest—were used to assess the effectiveness of the features employed in the in building the model.

Three criteria were established to help with the selection of machine learning algorithms for this study. First off, because machine learning models might be domain-specific, relevant literature on classification techniques for leaners engagement level was crucial in choosing a classifier. Second, recommendations were drawn from a survey of learners' engagement and performance prediction in e-learning studies to help with model choice. Thirdly, comparison results from large datasets had an impact on the choice of classification techniques. As a result, a variety of machine learning algorithms were evaluated employing WEKA.

3.4.1 Decision Tree

The Decision Tree algorithm works by categorizing data employing cycles of controls. Just as the name implies, the model has a structure of a tree consisting of nodes and leaves which provides a definite structure. The strength of this model lies in its ability to take care of attributes that are either unnecessary or appropriate [19]. The schooling process automatically chooses the utmost appropriate attributes and then produces the offspring nodes from the parent node [21].

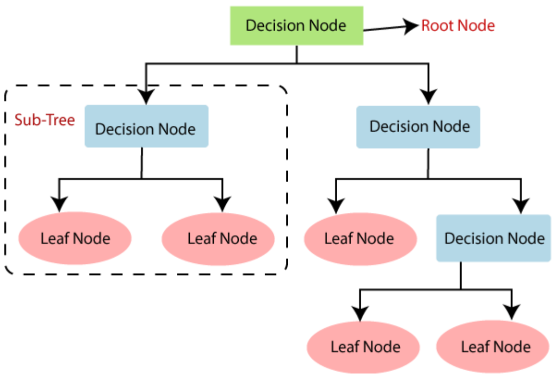

Figure 2. Decision Tree

3.4.2 Naïve Bayes

The Naïve Bayes algorithm has it foundation rooted in the Bayes theorem which was propounded by Thomas Bayes. The strength of this model lies in its ability to take care of missing values. Unlike other models, Naïve Bayes conserves both processing and schooling time [5]. The term 'naive' is used in describing this algorithm due to its hypothesis of uncertain independence. Hashim et al., [21] opined that if such an uncertain hypothesis really holds then Naïve Bayes sustains the ability to converge quicker when compared to several other models. The probability distribution can be written as

$$P\left(\frac{X}{Y}\right) = \frac{P\left(\frac{Y}{X}\right) * P(X)}{P(Y)} \qquad 3.1$$

Where X is the training set of attributes and Y is the given class.

3.4.3 Random Forest

Random Forest is a potent ensemble classifier built upon decision trees, amalgamates multiple such trees. This approach of merging classifiers confers distinctive characteristics upon the random forest, setting it apart from traditional tree classifiers. While a single decision tree classifier may be susceptible to disruptions from outliers or noise that impact overall model performance, the RF classifier introduces randomness to mitigate this issue [20]. Furthermore, the random forest introduces randomness not only to the data but also to the features. The random forest leverages principles akin to those found in bootstrapping and bagging classifiers, diversifying the trees by training them on distinct subsets of data created through bootstrap aggregation, regression and classification responsibilities appropriately [22].

$$Gini\ (D) = 1 - \sum_{i-1}^{c} P_i^2 \qquad 3.2$$

Where D is the set of instances in the node, c is the number of classes and P_i is the proportion of instances in class i in node D

3.4.4 Logistic Regression

Logistic regression is commonly employed in examining and describing the correlation that exists among entities with only two possible outcomes such as 'Yes' or 'No' and a sequence of foreseen entities [23]. Logistic Regression calculates the odds of several classes employing a boundary rationality distribution as depicted in the expression below

$$P = (Y = K|x) = \frac{e^{w_k * x}}{1 + \sum_K^{K-1} e^{w_k * x}} \qquad 3.4$$

Where k = 1, 2, 3…., k – 1 and x is the sample to be classified into the highest possible class.

3.4.5 Stochastic Gradient Descent

Stochastic Gradient Descent is a repetitive model that begins its journey from a non-specific locus and descends the slope in a stepwise manner aiming at the minimal point of the function thereby estimating the extent of variation of an entity with respect to the change of another entity [24]. The Stochastic Gradient Descent can be presented mathematically as

$$Q(w) = \frac{1}{n} \sum_{i-1}^{n} Q_i\ (w) \qquad 3.5$$

Where w is the constraint that reduces Q and is to be calculated. The individual aggregate function Q_i is related to the i^{th} performance in the set of data employed in training the model.

3.4.6 LogitBoost

Jerome Friedman articulated the improved algorithm which is an advanced version of the AdaBoost algorithm, which is employed for both binary and multi-class problems. This algorithm aims at reducing the logistic cost thereby improving performance [22]. When the logistic regression cost function is harnessed together with AdaBoost we derive the LogitBoost

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.41, September 2023 – www.ijais.org

algorithm [25].

$$F\ (x) = \sum_{i} \log(1 +\ e^{-y_i f(x_i)}) \qquad \textbf{3.6}$$

Where x_i signifies the value of the attribute and y_i signifies the class label.

3.4.7 Sequential Minimal Optimization (SMO)

SMO was discovered by John Platt in 1998. It is used to school the Supervised Vector Machine. SMO allows the unravelling and enhancement of quadratic problems by dividing the problem into two smaller chunks and then solving each iteratively. This algorithm has good management ability, especially in terms of memory. Furthermore, it is perceived to be the fastest when compared to other algorithms of the SVM family [21]

3.4.8 Voted Perceptron

Rosenblatt and Frank developed the perceptron framework. This framework works well with a set of data that are directly distinguishable in terms of their boundaries. This framework is employed due to ease of implementation and has proven to be more resourceful and less expensive with respect to the implementation period when compared to other algorithms like the Supervised Vector Machine. Researchers [5], [21], [26] have considered the Multi-layer Perceptron paying less attention to the Voted Perceptron, however, we see it necessary to examine the behaviour of this algorithm in predicting students' engagement.

3.4.9 Adaptive Boosting

Boosting is a collaborative modeling procedure founded in 1997 by Freund and Schapire. Over the years boosting is being employed in undertaking categorization problems involving two classes. Boosting framework aim at stepping up forecasting capability by changing ineffectual learners into effectual ones. Primarily there are three classes of boosting algorithms; Adaptive Boosting, Gradient Descent Boosting, and Extreme Gradient Descent Boosting algorithm. Boosters have the mathematical representation of the form;

$$F_T(x) = \sum_{t=1}^{T} f_t(x) \qquad \textbf{3.7}$$

where f_t is an ineffectual learner that acts on a data x, however, the t-th categorizer will be positive if the data being acted on are all in the positive class.

3.4.10 Support Vector Machine

Support vector machine is a binary linear. As a non-probabilistic supervised learning algorithm it leverages on training data an employs high dimensional space to generate a set of hyperplanes for classifying data. Although only test data features are provided, the model is built using training data to predict the target value. The superlative hyperplane must be chosen for SVM classification method to properly classify the problem instances.

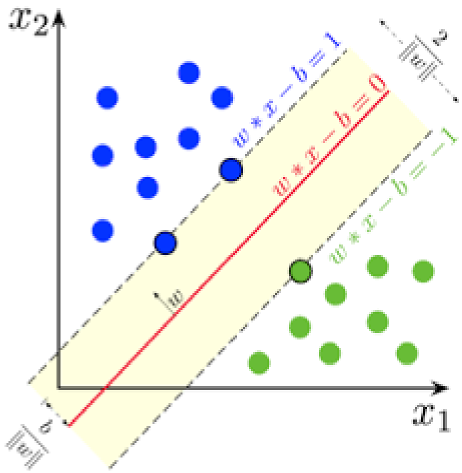

Figure 3. Support vector machine

4. EXPERIMENTAL SET-UP

The classification testing was embarked on to investigate leaners' engagement in a blended e-learning system. The investigation was carried out on a system running window 10 with a 64-bit operating system. The system uses an Intel Core™ i3 Core™ i3-380 CPU @ 2.53GHz with 4GB Random Access Memory (RAM). Waikato Environment for Knowledge Analysis (WEKA) version 3.8.6 was leveraged on to pre-process the data. We employed nine different machine learning models to evaluate learners' engagement level. The different models provide significant insight to their performance with respect to the used evaluation metrics. 10-fold cross-validation was employed in schooling the models. WEKA has different machine learning algorithms which runs on Java and provides a suitable environment for prediction and feature selection processes. We did not change the default settings. We employed precision, recall, accuracy and f-measure to access the performance of the models. The table below provides the parameter optimization and tuning values of the classifiers. Table 1 supplies the parameter optimization and tuning values of the classifiers employed in this research.

Table 1. Parameter Optimization and Tuning Values of the Classifiers

Classifier	Parameter	Value
Stochastic Gradient Descent	Batch size	100
	Seed	1
	Epochs	500
	Epsilon	0.001
AdaptiveBoost	Batch size	100
	Seed	1
	Weight threshold	100
LogitBoost	Batch size	100
	Seed	1
	Weight threshold	100
	Pool size	1
Decision Tree	Batch size	100
	Seed	1
	Confidence factor	0.25
Sequential Minimal Optimization	Batch size	100
	Seed	1
	Epsilon	1.0E-12
	Tolerance parameter	0.001

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.41, September 2023 – www.ijais.org

Voted Perceptron	Batch size	100
	Seed	1
	Exponent	1.0
	Iterations	1
Random Forest	Batch size	100
	Seed	1
	Max depth	0

4.1 Evaluation Metrics

Evaluation measures are metrics used to assess the results of an experiment. In the context of classification algorithms, different evaluation metrics are used to measure their output. In this study, the main performance evaluation metric employed is "Accuracy". However, additional metrics such as recall, precision, and f-measure are also used to supplement the evaluation of the framework's performance. Each classifier demonstrates its ability to identify learners' engagement when assessed using these metrics. A brief description of these metrics is provided below.

4.1.1 Accuracy (AC)

AC value is a widespread evaluation metric for classification models. It can be calculated as the ratio of well-predicted samples to the total sample of prediction. For a balanced dataset, accuracy is a fair indicator of the efficiency of the model Buraimoh [5] asserted.

$$AC = \frac{TP + TN}{TP + FP + FN + TN} \qquad 4.1$$

In this equation, a true number that is positive is denoted by TP while a true number that is negative is denoted by TN, however, FN denotes a false number that is negative and FP denotes a false positive number.

4.1.2 Precision (PR)

This metric is calculated by dividing all the true positive samples by the sum of the predicted positive samples and predicted negative samples.

$$PR = \frac{TP}{TP + FP} \qquad 4.2$$

A high precision score means a stronger model at predicting the classes, while a lower precision means a weak prediction of the classes [27].

4.1.3 Recall (RE)

RE calculates the correctly estimated positive results from all the observed positive results. We calculate the recall by dividing the true positive samples by the sum of the positive samples. This is sometimes called the detection rate.

$$RE = \frac{TP}{TP+TN} \qquad 4.3$$

4.1.4 F1-score (FS)

FS is vital as a bridge for recall and precision. It offers a measure of the findings that are incorrectly graded. It is regarded as the best metric for measuring the performance of models on an imbalanced dataset. Generally, the range of the FS ranges from 0 to 1. Higher values specify an excellent performance of the model.

$$FS = \frac{2PR*2RE}{PR+RE} \qquad 4.4$$

5. RESULTS AND DISCUSSION

This section supplies the outcome of the analysis employing bench-mark dataset. We run the classifiers using 10-fold cross validation. Accuracy is employed as the major metric for evaluating the performance of the models, nevertheless we supply the performance of the models using f-measure, recall and precision.

5.1 Experimental Results

Table 2 summarises the result of the 9 models after evaluating their performance using precision, accuracy, f-measure and recall. Figure 4 provides a visualization of the results. Figure 5, 6, 7, and 8 provides separate graphical representations of the individual evaluation metrics.

Classifiers	PRE (%)	RE (%)	FS (%)	AC (%)
SMO	89	89	89	90
SGD	89	89	89	90
RF	88	88	88	89
SLR	89	89	89	89
NB	89	89	89	88
LB	88	88	88	88
AB	88	88	88	88
DT	88	88	88	88
VP	73	79	78	79

From Table 2 it is observed that the performance of the models accuracy ranges from 90% and 79%. The result shows that all the models have the ability to predict learners' engagement level using both effort-related and interaction-related features. However, a careful analysis of the results shows that SMO and SGD provided the highest accuracy of 90%. With precision, recall, and f-measure value of 0.897 respectively SMO and SGD also outperformed the other models.

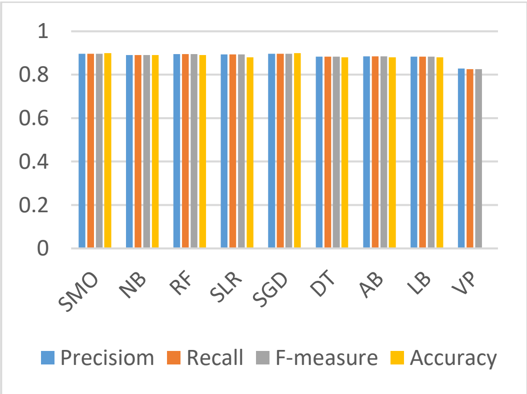

Figure 4: Graphical performance of different classification algorithms

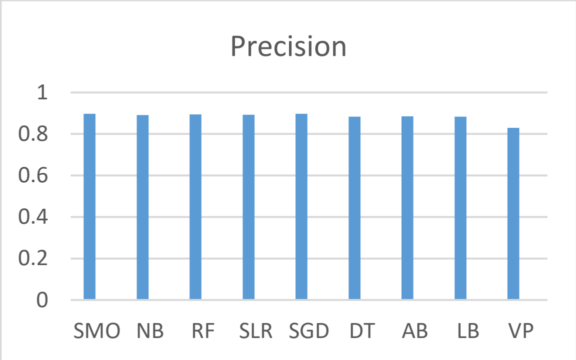

Figure 5: Graphical comparison of precision

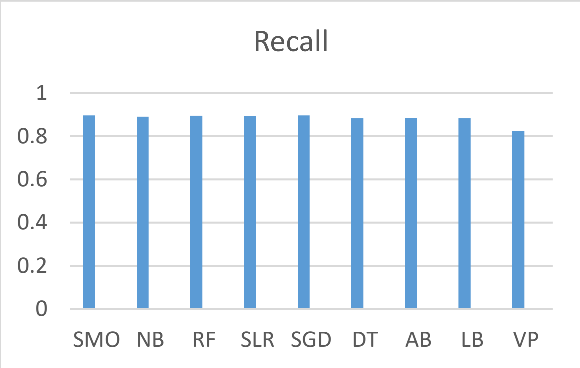

Figure 6: Graphical comparison of recall

It is important to note that although SMO and SGD provided the best accuracy, RF and SLR provide same accuracy of 89%, LB, AB, NB and DT provided same accuracy of 88% and VP provided the least accuracy of 79%. We therefore observe that SMO and SGD behave in a similar pattern with effort-related features and interaction-related features. RF and SLR have a similar pattern in analysing and predicting leaners' engagement and LB, AB, NB and DT also behave in a similar pattern in predicting engagement employing this features.

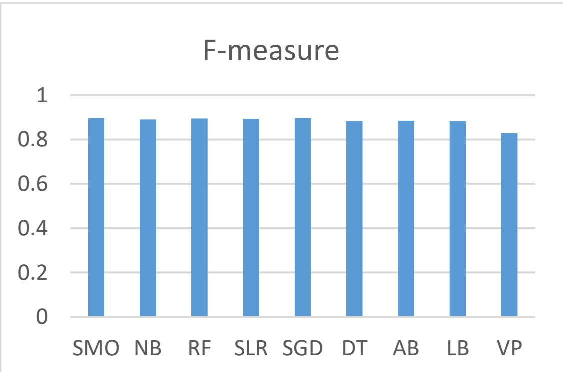

Figure 7: Graphical comparison of F-measure

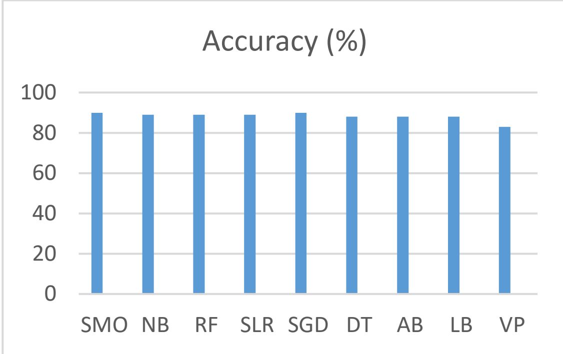

Figure 8: Graphical comparison of accuracy

The observation from the performance analysis reveals that the trained models have the ability to predict learners' engagement level in a blended e-learning system with an accuracy ranging from 70% to 90%.

5.2 Result Comparison Analysis with Existing Models

It is paramount to compare the outcome of this research with the outcome of previous researches in this domain, this will provide a good premise to ascertain the percentage of improvement achieved. After a careful comparison of the proposed model with three existing models we discovered significant improvement in the accuracy of the existing model. Table 3 provides a summary of the comparative analysis of these models.

Table 3: Result comparison with existing models

Author	AC (%)	PR (%)	RC (%)	FS (%)
Hussain et., al. [10]	76	-----	93	-----
Gorgun et., al. [12]	66	66	66	65
Jawad et., al. [13]	82	-----	85	89
Proposed Model	**90**	**89**	**89**	**89**

From Table 2, we present the accurate performance of the proposed model against the performance existing models. We observed that the proposed model recorded 90% accuracy which is higher when compared to 66%, 76% and 82% recorded by previous studies. This proves the efficiency of the proposed model. A closer comparison reveals that the proposed model outperforms the existing models by 8%, 24% and 16% respectively. However the difference between the proposed model and the least performed existing model is 24%.

As revealed by Table 2, the improved model has recorded a significant level of enhancement especially when compared to the model developed by Gorgun et al., [12], while they recorded an accuracy of 66%, the improved model in this research recorded an accuracy of 90%. This level of accuracy is commendable for a predictive model. When compared the model proposed by Jawad et al., [13] we recorded 14% improved which is very significant.

6. CONCLUSION AND FUTURE WORK

Students' engagement level prediction has been a significant challenge in e-learning. Learners' engagement level prediction is a task that distinguishes learners with low degree of involvement from learners with high degree of involvement. Accurate identification of low engaged learners' will help in reducing the dropout rate associated with distance learning. To overcome the limitations associated with learners engagement, this study leverages benchmark dataset and employs various classifiers (SMO, NB, RF, SLR, SGD, DT, AB, LB and VP) to build a machine learning model for leaners engagement level prediction in e-learning. An extensive analysis was carried to access the performance of the models. Information Gain feature ranker was employed in accessing the significance of the features employed. Accuracy was used as the major performance measure due to its ability to provide exact corrections of models performance. SMO provided the highest accuracy of 90% with the highest precision (0.897), recall (0.897) and f-measure (0.897). Comparing the highest performance of the trained model with baseline models reveals the significance of employing Information Gain feature ranking in evaluating the significance of the features. The analysis of the features reveals that effort-related and interaction-related features are significant in predicting learners' engagement level in a blended e-learning system.

In the future the researchers intend to extend the scope of this research by building a hybrid prediction model with the ability to forecast leaners engagement level both in a blended and online e-learning platform. The researchers are also considering building a deep learning model that will provide significant insight into the learners' engagement level.

7. STATEMENTS AND DECLARATIONS

Conflict of Interest Declaration: The authors declare that there is no conflict of interest.

8. REFERENCES

[1] A. Giannoulas, A. Stampoltzis, K. Kounenou, and A. Kalamatianos, "How Greek Students Experienced Online Education during Covid-19 Pandemic in Order to Adjust to a Post-Lockdown Period," vol. 19, no. 4, pp. 222–232, 2021.

[2] R. Karasneh, S. Al-azzam, S. Muflih, S. Hawamdeh, and Y. Khader, "Attitudes and Practices of Educators Towards e-Learning During the COVID-19 Pandemic," vol. 19, no. 4, pp. 252–261, 2021.

[3] A. M. Nortvig, A. K. Petersen, and S. H. Balle, "A literature review of the factors influencing e-learning and blended learning in relation to learning outcome, student satisfaction and engagement," *Electron. J. e-Learning*, vol. 16, no. 1, pp. 45–55, 2018.

[4] R. Khandelwal and U. Kumar, "Available on : SSRN Applications of Artificial Neural Networks in E-Learning Personalization," 2020.

[5] E. F. Buraimoh, "Predicting Student Success Using Student Engagement in the Online Component of a Blended-Learning Course," pp. 1–65, 2021.

[6] T. Anderson, "Applications Of Machine Learning To Student Grade Prediction In Quantitative Business Courses," vol. 1, no. 3, pp. 13–22, 2017.

[7] F. Junshuai, "Predicting Students' Academic Performance with Decision and Neural Network," *Αγαη*, vol. 8, no. 5, p. 55, 2019.

[8] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10031042.

[9] S. Leelavathy, R. Jaichandran, S. S. K, B. Surendar, A. K. Philip, and D. R. Ravindra, "Students Attention and Engagement Prediction Using Machine Learning Techniques," vol. 7, no. 4, pp. 3011–3017, 2020.

[10] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/6347186.

[11] G. Okereke, "A Machine Learning Based Framework for Predicting Student's Academic Performance," *Phys. Sci. Biophys. J.*, vol. 4, no. 2, 2020, doi: 10.23880/psbj-16000145.

[12] G. Gorgun, S. N. Yildirim-Erbasli, and C. Demmans Epp, "Predicting Cognitive Engagement in Online Course Discussion Forums," no. July, pp. 276–289, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.6853149.

[13] K. Jawad, M. A. Shah, and M. Tahir, "Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing," *Sustain.*, vol. 14, no. 22, 2022, doi: 10.3390/su142214795.

[14] N. Alruwais and M. Zakariah, "Student-Engagement Detection in Classroom Using Machine Learning Algorithm," *Electron.*, vol. 12, no. 3, 2023, doi: 10.3390/electronics12030731.

[15] R. S. Olson, W. La Cava, P. Orzechowski, R. J. Urbanowicz, and J. H. Moore, "PMLB: A large benchmark suite for machine learning evaluation and comparison," *BioData Min.*, vol. 10, no. 1, pp. 1–13, 2017, doi: 10.1186/s13040-017-0154-4.

[16] M. R. Segal, "Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression," *Cent. Bioinforma. Mol. Biostat.*, p. 15, 2004, [Online]. Available: https://escholarship.org/uc/item/35x3v9t4.

[17] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student Engagement Level in an e-Learning Environment: Clustering Using K-means," *Am. J. Distance Educ.*, vol. 34, no. 2, pp. 137–156, 2020, doi: 10.1080/08923647.2020.1696140.

[18] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions," *IEEE Access*, vol. 7, pp. 144907–144924, 2019, doi: 10.1109/ACCESS.2019.2944243.

[19] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *Appl. Sci.*, vol. 9, no. 20, 2019, doi: 10.3390/app9204396.

[20] C. I. Eke, A. A. Norman, and L. Shuib, *Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach*, vol. 16, no. 6

June. 2021.

[21] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 928, no. 3, 2020, doi: 10.1088/1757-899X/928/3/032019.

[22] H. Hassan, N. B. Ahmad, and S. Anuar, "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," *J. Phys. Conf. Ser.*, vol. 1529, no. 5, 2020, doi: 10.1088/1742-6596/1529/5/052041.

[23] M. B. Yildiz and C. Borekci, "Predicting Academic Achievement with Machine Learning Algorithms," vol. 3, no. 3, 2020, doi: 10.31681/jetol.773206.

[24] B. K. Francis and S. S. Babu, "Predicting Academic Performance of Students Using a Hybrid Data Mining Approach," *J. Med. Syst.*, vol. 43, no. 6, 2019, doi: 10.1007/s10916-019-1295-4.

[25] P. Sun, M. D. Reid, and J. Zhou, "An improved multiclass LogitBoost using Adaptive-One-Vs-One," *Mach. Learn.*, vol. 97, no. 3, pp. 295–326, 2014, doi: 10.1007/s10994-014-5434-3.

[26] D. N. Lu, H. Q. Le, and T. H. Vu, "The factors affecting acceptance of e-learning: A machine learning algorithm approach," *Educ. Sci.*, vol. 10, no. 10, pp. 1–13, 2020, doi: 10.3390/educsci10100270.

[27] A. K. H. Ali Salah Hashim, Wid Akeel Awadh, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms Student Performance Prediction Model based on Supervised Machine Learning Algorithms," 2020, doi: 10.1088/1757-899X/928/3/032019.