**Pneumonia Image Classification: A Comparative Study of Naive Bayes, Support Vector Machine, and Logistic Regression Models**

**Mike Rasell Carale Dago-oc1, Chris Angelu Bongcawil Jordan2**

Computer Science and Information Technology Department,

Bachelor of Science in Computer Science (BSCS)

Negros Oriental State University (NORSU), Negros Oriental, Philippines

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***  Received November 14, 2024  Revised January 10, 2025  Accepted January 10, 2025 |  | Pneumonia is a common and fatal disease, which is also one of the leading causes of death worldwide. One of the ways to diagnose this disease is through chest X-ray imaging, and with this, there is a need for an expert radiologist who possesses expertise and experience in the desired domain. According to the World Health Organization (WHO) report, about 2/3 of people in the world still do not have access to a radiologist to diagnose their disease. With this problem in mind, this study proposes a machine-learning classification model that diagnoses pneumonia disease efficiently and effectively. The study used a dataset of 290 X-ray images with 145 normal and 145 pneumonia images. The machine learning algorithms used are: Naive Bayes, Support Vector Machine, and Logistic Regression. The performance metrics of each model were done using 10-fold cross-validation. The predictive results show that the Support Vector Machine outperformed other models by attaining an accuracy of 94.83% with precision, recall, and f1-score values of 0.9528, 0.9483, and 0.9480 respectively. LR and NB also demonstrated strong performance, with LR having an accuracy of 94.14% while NB has an accuracy of 91.03%.  *This is an open access article under the* [*CC BY-SA*](https://creativecommons.org/licenses/by-sa/4.0/) *license.* |
| ***Keywords:***  Pneumonia  Naïve Bayes  Support Vector Machine (SVM)  Logistic Regression  Inception-V3  Accuracy  Precision  Recall  F1-Score |

1. **INTRODUCTION**

Pneumonia is one of the leading causes of death worldwide [1], it can affect everyone anywhere, especially children [2]. Pneumonia is a form of acute respiratory infection that is most commonly caused by viruses or bacteria targeting the lungs and it can cause mild to life-threatening illness in people of all ages. When an individual has pneumonia, the alveoli are filled with pus and fluid, which makes breathing painful and limits oxygen intake [3]. Understanding student engagement in academic institutions provides crucial insights into the effectiveness of teaching methods. This data serves as a powerful tool for educators and researchers to refine strategies that enhance student learning. It provides objective insights into student experiences, going beyond assumptions and anecdotes to reveal their true engagement. This valuable information helps institutions improve academic programs while also supporting marketing, recruitment, and addressing students' evolving learning needs [4]. Machine Learning (ML) is being employed by researchers to analyze this data and gain a deeper understanding of the patterns within it [5].

Identifying if a person has pneumonia requires general practitioners or doctors specializing in lung conditions [4]. Sadly some countries have not kept pace with the overall economic growth, and their health systems remain weak, which includes doctors who specialize in lung conditions [5]. This lack of general practitioners leads to increased workloads for existing specialists, potential delays in diagnosis and treatment, and overall strain on healthcare systems [6].

This study aims to make a machine-learning model that can classify X-ray images if it has pneumonia. With this tool, it can potentially automate diagnosis which can reduce the time doctors spend on diagnosing pneumonia cases. This allows doctors to focus more on treatment rather than performing or interpreting tests. Additionally, this can also be used as a training and education tool for medical students to learn how to interpret chest X-rays. However, the model is intended solely as a support tool, not a replacement for human judgment.

The rest of this paper is structured thus: Section 2 provides a detailed review of recent literature in the field of pneumonia image classification, Section 3 outlines the methodology employed in this study, and further provides a brief description of the machine learning algorithms used in this investigation. While Section 4 presents the results and discussion, Section 5 concludes the research, and Section 6 offers recommendations.

1. **LITERATURE REVIEW**

Machine learning (ML), once a theoretical concept, has rapidly evolved into a transformative force shaping nearly every aspect of our lives [7]. This technology is also slowly integrated into healthcare, offering innovative solutions for diagnosis, treatment, and operational efficiency [8]. For instance, ML models are being utilized to detect pneumonia from chest X-rays [9]. This section is a summary of existing related papers previously published about detecting pneumonia from chest X-rays that also use Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR).

The researchers in [9] tested the predictive performance of 6 machine learning models. In this review, we focus on Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR). The other models include Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), and AdaBoost (AB). Their dataset consists of 5,856 X-ray images, which include 1,583 normal cases and 4,273 pneumonia cases and is split into 70% for training and 30% for testing. For image embedding they used 5 Deep Convolutional Neural Networks (DCNN) as feature extractors. However, this study will only focus on Inception-V3, the same as the one used in this study. The other DCNNs are AlexNet, SqueezeNet, VGG16, and VGG19. ANN delivered the best overall performance with a 97.19% accuracy, closely followed by LR at 97.08%. Meanwhile, SVM with a linear kernel achieved an accuracy of 85.02%.

Researchers in [10] used five models including, Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Softmax. In image embedding they extracted the features from the fully connected (FC) layer of a modified CNN architecture using transfer learning. The FC layer generated feature vectors representing each class. The CNN model used in this study was pre-trained on Optical Coherence Tomography (OCT) images before being adapted to the pneumonia detection task. This differs from directly using models like Inception V3. Their dataset is comprised of 5,852 anterior-posterior chest X-ray images, which include 1,581 normal and 4,271 pneumonia cases, and is split into 70% for training, 15% for validation, and 15% for testing. However, in this review, we focus on the SVM, which achieved an accuracy of 99.61%. Among the models, Softmax slightly outperformed SVM with an accuracy of 99.72%.

While Researchers in [11] employed several models, including Multilayer Perceptron (MLP), Random Forest (RF), Sequential Minimal Optimization (SMO), Logistic Regression (LR), and Classification via Regression. The research focused on pixels in lungs segmented Region of Interest (ROI) that are more contributing toward pneumonia detection than the surrounding regions, thus the features of lungs segmented ROI confined area are extracted. It also utilized a total of 412 chest X-ray images containing 206 normal and 206 pneumonic cases from the ChestX-ray14 dataset. But this study will only focus on the LR model which got an accuracy of 95.63% which is the highest among the five.

1. **METHODOLOGY**

This section provides an outline of the research methodology employed in this study.

**3.1. Hardware and Software**

The study was carried out on a system running Windows 10 with a 64-bit operating system. The system uses an Intel Core™ i5-7400 CPU with 16GB Random Access Memory (RAM). The researchers utilized Jupyter and Python version 3.9.15 as the primary programming language for data analysis and model implementation with the following libraries: Math, NumPy, Pandas, PIL, OS, Scikit-Learn, Matplotlib, Seaborn, TensorFlow, and Keras.

**3.2. Data Acquisition**

This research employs a dataset from Kaggle [12], a file with 5856 .jpeg files which are anterior-posterior chest X-ray images. The images are carefully chosen from retrospective pediatric patients with age groups between 1 and 5 years from Guangzhou Women and Children’s Medical Center. The image has 2 normal categories pneumonia and normal which are further divided for testing, training, and evaluation [12]. Because of hardware limitations, this study uses 290 images with 145 normal and 145 pneumonia images.

**3.3. Image Embedding**

The main goal and objectives of the proposed system are to diagnose and make a tool that can identify whether a person has pneumonia through chest X-ray images. This study uses a Deep Convolutional Neural Network (DCNN) which is Inception-V3, for feature extraction of the X-ray images.

Inception models were developed by a researcher [13] for the first time in 2014. The structures of inception models and the conventional CNN model are different because they are inception blocks which means lapping the same input tensor with multiple filters and concatenating their results. In 2015, a researcher [15] proposed a new version of the inception models named Inception-V3, an improved version of the previous versions of inception models which are Inception-V1 and Inception-V2, and possesses 24M parameters. The advancement in Inception-V3 was as follows: first, it factorizes the “n × n” convolution into asymmetric convolutions which are 1 × n and n × 1. Second, it factorizes the 5 × 5 convolutions into two 3 × 3 convolutions and lastly, it replaces 7 × 7 convolutions to a series of 3 × 3 convolutions. It consists of a block of convolutional layers arranged in a parallel manner and each layer consists of different sizes of filters 1 × 1, 3 × 3, and 5 × 5, respectively. Furthermore, 3 × 3 max pooling is also performed. The outputs are concatenated and sent to the next inception module [9].

**3.4. Data Pre-processing**

Data processing is an essential aspect of model development. Data acquired in their raw form contain noise and anomalies, which can affect the performance and training process of the model being schooled [15]. In addition to this when preparing input data for a DCNN, it's essential to adhere to specific requirements regarding input shape, color channels, and preprocessing steps to ensure optimal performance [16]. The researchers employed several techniques to clean the data, which included image pre-processing and data normalization:

Image pre-processing is an essential part of image classification to ensure that input images align with the model's expectations. Inception-V3 requires input images in 299x299 resolution and should also have three color channels (red, green, and blue) [17][18]. And since the datasets used are X-ray images they only have 1 color channel (black and white) and inconsistent high resolution. With this in mind, the images are turned to have three color channels, and it is also resized to have a resolution of 299x299. After this, the image is preprocessed using TensorFlow Keras API preprocess\_input so that the images align with the model's training conditions.

Data normalization is a pre-processing technique primarily intended to manage numerical features and is applied to numerical features before the application of classification algorithms. Normalization is crucial to prevent the effect of certain features from being concealed by others, particularly when the ranges of the features are inconsistent [19]. After feature extraction using Inception-V3, the features are normalized to improve model training [20]. The normalization techniques used are Min-Max Normalization for Logistic Regression (LR) and Naive Bayes (NB), and Z-Score Normalization for the Support Vector Machine (SVM).

**3.5. Machine Learning Algorithms**

This section focuses on the machine learning classification models utilized in this study. After image and data pre-processing, the machine learning workflow progresses to the model training stage, where an algorithm is taught to learn from data and produce predictions. This algorithm is specifically responsible for the classification of whether an X-ray image has pneumonia. To find the best classifier for engagement level prediction, several classifiers, including Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR) were tested through a variety of tests.

**3.5.1. Naïve Bayes**

The Naïve Bayes algorithm has its foundation rooted in the Bayes theorem by Thomas Bayes. One of the strengths of this model is its ability to handle missing values. And Unlike other models, Naïve Bayes conserves processing and training time [21][7]. The term ‘naive’ is used due to this algorithm's uncertain independence. With this, researchers in [22] stated that with this ability it's able to converge quicker when compared to several others.

(1)

Where X is the training set of attributes and Y is the given class [7].

**3.5.2. Support Vector Machine**

Support Vector Machine (SVM) is a binary linear classifier. As a non-probabilistic supervised learning algorithm, it utilizes training data and employs a high-dimensional space to construct a set of hyperplanes for data classification. While only the features of test data are provided, the model is trained on the training data to predict the target values. For effective classification of problem instances, SVM relies on selecting the optimal hyperplane [7].

**3.5.3. Logistic Regression**

Logistic regression is often used to analyze and describe the relationship between entities when the outcome has only two possible options, such as 'Yes' or 'No,' and a set of predictive entities [23]. Logistic Regression calculates the odds of several classes employing a boundary rationality distribution as depicted in the expression below [7].

(2)

**3.5 Model Evaluation Metrics:**

Evaluation measures are metrics used to assess the results of an experiment [24]. In the context of classification models, different evaluation metrics are used to measure their output. In this study, the main performance evaluation metric is “Accuracy”. However, additional metrics such as recall, precision, f-measure, and confusion matrices are also used to supplement the evaluation of the model's performance. Each model identifies learner engagement levels when assessed using these metrics. A brief description of these metrics is provided below.

**Accuracy (AC)** is a common evaluation metric for classification models. It's calculated as the ratio of well-predicted samples to the total sample of prediction. For a balanced dataset, accuracy is a reliable measure of the model's performance [22].

(3)

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 [7].

**Precision (PRE)** is computed by dividing all the true positive samples by the sum of the predicted positive samples and predicted negative samples [7].

(4)

A high precision score indicates strong class predictions, while a low precision score reflects weak class predictions [25].

**Recall (RE)** is the ratio of correctly predicted positive results to all actual positive samples, also known as the detection rate. It's calculated by dividing the true positive samples by the sum of the positive samples [7].

(5)

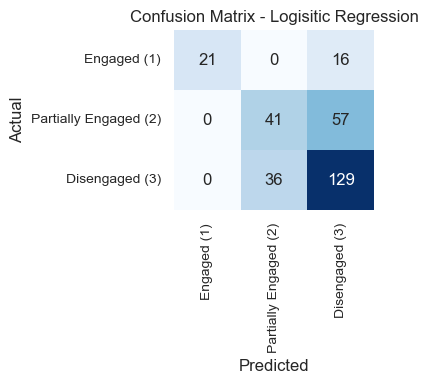
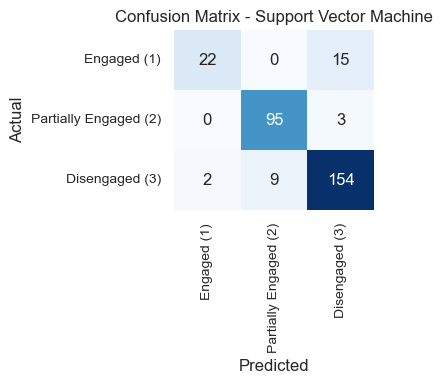
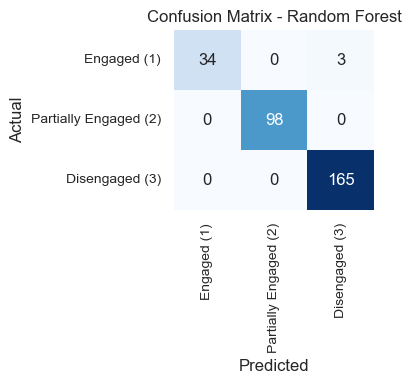
**F1-score (FS)** is the mean value for recall and precision. It offers an indicator of mistakenly graded results [22]. It is regarded as the best metric for measuring the performance of models on an imbalanced dataset. It ranges from 0 to 1, with higher values indicating better model performance [7].

(6)

1. **RESULTS AND DISCUSSION**
2. **RESULTS AND DISCUSSION**

This section provides the outcome of the analysis, achieved through running the models using 10-fold cross-validation, with accuracy as the primary metric for evaluating the performance of the models. Additionally, the performance of the models is further assessed using f1-score, recall, precision, and confusion matrices.

**4.1. Model Performance**

****Figure 1illustrates the performance of three machine learning models which are RF, SVM, and LR in classifying student engagement levels into three categories: Engaged (1), Partially Engaged (2), and Disengaged (3). The results of the three models after evaluating their performance using precision, accuracy, f1-score, and recall are in Table 1, while Figure 2 provides a visualization of the results.

**Figure 1.** Confusion Matrices

RF performed very well, especially in classifying the Partially Engaged (2) and Disengaged (3) classes. Although it misclassified a small number of Engaged (1) samples, RF strong mechanism allowed it to effectively know the patterns of the data, leading to the high performance of the model.

SVM also performs significantly better than Logistic Regression but slightly worse than Random Forest. It performed very well particularly in Partially Engaged (2) and Disengaged (3) with minor errors. There are still misclassifications between Engaged (1) and Disengaged (3), similar to LR. Overall, the model shows strong predictive performance across all three classes compared to LR.

In contrast, LR struggles with the classification task. LR struggles to classify the Engagement Level classes accurately, particularly for Partially Engaged (2) and Disengaged (3) classes. Engaged (1) is partially correct but still has significant misclassifications. The model frequently predicts Disengaged (3) even for samples that are Engaged (1) or Partially Engaged (2). The confusion between classes indicates that a linear model like LR cannot capture the complex relationships in the data.

Overall, RF is the most effective model in predicting engagement levels, followed by SVM, while LR performs inadequately in this context.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-Score |
| Random Forest (RF) | 0.9729 | 0.9709 | 0.9729 | 0.9709 |
| SVM | 0.9100 | 0.9038 | 0.9100 | 0.9156 |
| Logistic Regression (LR) | 0.6943 | 0.6853 | 0.6943 | 0.7061 |

**Table 1**. Performance Metrics

From Table 1 it can be observed that the performance of all the models accuracy ranges from 69% to 97%. The result shows that all the models can predict learner engagement levels using biometrics and environmental data. However, a thorough analysis of the results shows that RF provided the highest accuracy of 97.29%. With a precision and F1-score of 97.09%, and a recall of 97.29%, same as its accuracy. This suggests that RF handles the dataset very well, which shows that the dataset have non-linear relationships and mixed features. SVM also performs very well with accuracy close to RF. Using kernel functions like the Radial Basis Function (RBF) shows that SVM is a powerful model for non-linear classification. While LR struggles with the dataset because it assumes a linear relationship between the input features and the target.

**Figure 2.** Graphical Performance Metrics

Figure 2 presents the performance metrics through a line chart which are: accuracy, precision, recall, and F1-score of RF, SVM, and LR.

RF achieves the highest values across all metrics, with performance around 97–98%. The minimal variance between metrics indicates RF's overall reliability and balanced performance.

SVM maintains strong performance across all metrics, with values close to 91%. While slightly lower than RF, SVM proves to be an effective classifier for this dataset.

LR exhibits significantly lower performance, with metrics around 69–70%. This highlights its limitations in modeling complex data, as it struggles to accurately classify samples. The gap between LR and the other models is considerable, indicating its unsuitability for this dataset.

The graph demonstrates that RF outperforms SVM and LR in all performance metrics, making it the best model for predicting engagement levels. SVM, while slightly less effective than RF, still provides strong results and is a viable alternative. LR, on the other hand, performs poorly due to its inability to capture the complexity of the data, making it the least effective model. For predicting engagement levels, RF or SVM should be preferred over LR.

**4.2. Comparative Discussion**

The researchers in this study have evaluated the performances of the three models namely, Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) in classifying student engagement based on their biometric and environmental data. The models demonstrate varied strengths and weaknesses, highlighting their effectiveness in classification tasks.

RF has been shown to outperform the two other models, with consistent accuracy, precision, recall, and f1-score of 97%. The model’s ability to handle non-linear relationships and complex interactions between features made it an effective model for this study. Also, the confusion matrix shows perfect accuracy in predicting partially engaged (2) and disengaged (3) classes.

LR performs poorly compared to RF and SVM due to its inability to capture complex, non-linear relationships in the data, leading to higher misclassification rates, especially for overlapping classes like Partially Engaged (1) and Disengaged (3) classes.

The SVM model performs significantly better than LR and is slightly behind RF. Its ability to effectively separate the classes (Engaged, Partially Engaged, and Disengaged) shows its strength in handling the complexity of the dataset. While SVM performs well, RF maintains a slight edge with higher precision and accuracy, making it the top-performing model overall.

In this study, the performances of RF, LR, and SVM were evaluated for classifying student engagement based on biometric and environmental data. And among the three RF is the most effective model for classifying student engagement levels, followed by SVM, while LR struggles with the complexity of the data.

1. **CONCLUSION**

Real-time student engagement prediction has been a significant challenge in many school institutions. This prediction classifies students based on their degree of involvement or engagement and can be either, engaged, partially engaged, or disengaged. Accurate real-time prediction of student engagement can overcome lower achievement and high dropout rates in schools. To overcome the limitations associated with traditional student engagement classifiers, this study uses biometric and environmental data together with three classifiers, which are Random Forest, Logistic Regression, and Support Vector Machine, to build a machine learning algorithm that can predict student engagement levels. After a thorough analysis of the performance of each model, RF provided the highest accuracy, precision, recall, and f1-score of 97%. The findings of this study suggest that RF is the most effective model for predicting student engagement, offering a promising tool for institutions aiming to identify and address engagement issues early and in real time, while SVM and Logistic Regression provide alternative solutions with varying levels of performance.

1. **RECOMMENDATION**

The researchers in the study suggest several recommendations to improve the scope and application of this study. Other machine learning models can be explored like neural networks and gradient boosting to see if they can outperform RF. Additionally, ongoing model evaluation using real-time data will ensure the model's effectiveness in different school settings. And while LR performed poorly in this study, it might still be useful in less complex scenarios. If simplicity and interpretability are required, LR could be employed with some improvements, such as feature engineering or combining it with more advanced models in an ensemble approach. Lastly, SVM performed better than LR and could serve as an alternative when RF is not feasible due to computational constraints or the need for simpler models. Further experimentation with kernel types and parameter tuning could enhance its performance for this specific task.

**REFERENCES:**

[1] NWEA, "Research proof points: Better student engagement improves student learning," Teach. Learn. Grow., 2015. Available: <https://www.nwea.org/blog/2015/research-proof-points-better-student-engagement-improves-student-learning/>.

[2] [1] T. D. Nguyen, M. Cannata, and J. Miller, "Understanding student behavioral engagement: Importance of student interaction with peers and teachers," 2016. Available: <https://files.eric.ed.gov/fulltext/ED578739.pdf>.

[3] [1] J. S. Renzulli, "The relationship between student engagement and student achievement," University of Connecticut, 2024. Available: <https://gifted.media.uconn.edu/wp-content/uploads/sites/961/2024/04/The-Relationship-Between-Student-Engagement-and-Student-Achievement.pdf>.

[4] A. P. Delfino, "Student engagement and academic performance of students of Partido State University," 2019. Available: <https://files.eric.ed.gov/fulltext/EJ1222588.pdf>.

[5] [1] G. Theophilus and C. I. Eke, "Machine learning-based e-learners' engagement level prediction using benchmark datasets," Int. J. Appl. Inf. Syst., vol. 12, no. 41, Sept. 2023.

[6] T. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," Int. J. Machine Learn., vol. 1, no. 3, pp. 13–22, 2017.

[7] T. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," Int. J. Appl. Inf. Syst., vol. 1, no. 3, pp. 13–22, 2017.

[8] S. Ayouni, F. Hajjej, M. Maddeh, and S. Al-Otaibi, "A new ML-based approach to enhance student engagement in online environment," PLOS ONE, Nov. 10, 2021. Available: <https://doi.org/10.1371/journal.pone.0258788>.

[9] N. Alruwais and M. Zakariah, "Student-engagement detection in classroom using machine learning algorithm," MDPI Electronics, vol. 12, no. 3, p. 731, 2023. Available: <https://www.mdpi.com/2079-9292/12/3/731>.

[10] 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/psbj16000145.

[11] M. A. Hernández-Mustieles, Y. E. Lima-Carmona, M. A. Pacheco-Ramírez, A. A. Mendoza-Armenta, J. E. Romero-Gómez, C. F. Cruz-Gómez, D. C. Rodríguez-Alvarado, A. Arceo, J. G. Cruz-Garza, M. A. Ramírez-Moreno, and J. de J. Lozoya-Santos, "Wearable biosensor technology in education: A systematic review," Sensors, vol. 24, no. 8, pp. 2437, 2024.

[12] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," Artificial Intelligence Review, vol. 52, no. 2, pp. 987–1011, 2019, doi: 10.1007/s10462-018-9620-8.

[13] J. McKee, "Experimental research: A quantitative research method," CJ Docs Research Glossary, 2024. Available: <https://docmckee.com/cj/docs-research-glossary/experimental-research-a-quantitative-research-method/>.

[14] Z. Ziya, "Emotional monitoring dataset," Kaggle, 2023. Available: <https://www.kaggle.com/datasets/ziya07/emotional-monitoring-dataset>.

[15] 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.

[16] E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. M. F. El-Salhi, "The effect of preprocessing techniques, applied to numeric features, on classification algorithms’ performance," Data, 2021, 6(2), 11. <https://doi.org/10.3390/data6020011>.

[17] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed.; Morgan Kaufmann: San Mateo, CA, USA, 2011.

[18] Edvancer, "Logistic Regression vs Decision Trees vs SVM (Part 2)," Edvancer, 2021. Available: <https://edvancer.in/logistic-regression-vs-decision-trees-vs-svm-part2/>.

[19] PureAI, "Machine learning techniques in AI applications," PureAI, Apr. 10, 2020. Available: <https://pureai.com/articles/2020/04/10/ml-techniques.aspx>.

[20] Javatpoint, "Performance metrics in machine learning," Javatpoint, 2021.. Available: <https://www.javatpoint.com/performance-metrics-in-machine-learning>.

[21] E. F. Buraimoh, "Predicting student success using student engagement in the online component of a blended-learning course," M.Sc. Research Report, School of Computer Science and Applied Mathematics, Faculty of Science, The University of the Witwatersrand, Johannesburg, South Africa, May 20, 2021. Available: <https://wiredspace.wits.ac.za/server/api/core/bitstreams/fb368595-69f9-4421-89d3-fc454199b9a0/content>.

[22] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student performance prediction model based on supervised machine learning algorithms," IOP Conf. Series: Materials Science and Engineering, vol. 928, no. 3, 032019, 2020. doi:10.1088/1757-899X/928/3/032019.

[23] C. I. Eke, A. A. Norman, and L. Shuib, "Multi-feature fusion framework for sarcasm identification on Twitter data: A machine learning-based approach," PLOS ONE, vol. 16, no. 6, p. e0252918, Jun. 10, 2021. <https://doi.org/10.1371/journal.pone.0252918>.

[24] 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, p. 052041, 2020. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1529/5/052041/pdf>.

[25] M. B. Yıldız and C. Börekçi, "Predicting academic achievement with machine learning algorithms," J. Educ. Technol. Online Learn., vol. 3, no. 3, pp. 372–392, 2020. Available: <https://files.eric.ed.gov/fulltext/EJ1294138.pdf>.