ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL
Available Online: https://assajournal.com
Vol. 05 No. 02. April-June 2026.Page# 1361-1370
Print ISSN: 3006-2497 Online ISSN: 3006-2500
Platform & Workflow by: Open Journal Systems
https://doi.org/10.5281/zenodo.20458105

A Comprehensive Review of Machine Learning Algorithms for Predicting Student Dropouts in Educational Data Mining

Rabia Bhutto
Department of Software Engineering, Mehran University of Engineering and Technology Jamshoro, Pakistan
rabeeabhutto99@gmail.com
Kanwal Batool
Department of Computer Systems Engineering Mehran University of Engineering and Technology Jamshoro, Pakistan
kanwalbatool149@gmail.com
Pireh Soomro
Department of Software Engineering, Mehran University of Engineering and Technology Jamshoro, Pakistan
pirehsoomro3@gmail.com
Muhammad Hasnain Ali Dad
Artificial Intelligence, Mehran University of Engineering and Technology Jamshoro, Pakistan
mhadadsoomro@gmail.com
Muhammad Noor Murtaza Qureshi
Artificial Intelligence MUET, Jamshoro
an3phemeral@proton.me

ABSTRACT

Student dropout is one of the major challenges faced by educational institutions, as it negatively affects aca-demic performance, institutional reputation, and overall stu-dent success. This study presents a machine learning-based framework for predicting student dropout using educational data mining techniques. A publicly available dataset obtained from the UCI Machine Learning Repository was utilized for experimental analysis. The dataset contains students' demo-graphic information, academic performance, attendance records, assignment results, socio-economic background, and engagement-related attributes. Several machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were implemented and evaluated for student dropout prediction. Data preprocessing techniques such as cleaning, normalization, and handling of imbalanced class distributions were applied to improve model reliability and prediction performance. The dataset was divided into training and testing subsets, while 10-Fold Cross-Validation was used to ensure robustness and generalization capability of the models. The performance of the classifiers was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Experimental results demonstrated that the Support Vector Machine (SVM) achieved the highest classification performance with an accuracy of 93.80%, precision of 0.939, recall of 0.938, and F1-score of 0.938, significantly outperforming the remaining classifiers. Random Forest, Multilayer Perceptron, KNN, Decision Tree, and Logistic Regression achieved comparatively lower prediction accuracies. The findings further revealed that academic performance, atten-dance percentage, assignment scores, engagement level, and socio-economic background are among the most influential factors affecting student dropout prediction. The proposed machine learning

framework provides an efficient and practical solution for identifying at-risk students at an early stage, enabling educational institutions to implement timely intervention strategies to improve student retention and academic success.
***Keywords**: Educational data mining, Machine Learning, Student dropout, Prediction, Classification*

1. Introduction

Student dropout is considered one of the major challenges faced by educational institutions worldwide because it affects academic performance, institutional reputation, and overall educational development. Educational Data Mining (EDM) has emerged as an important research area that helps in extracting meaningful knowledge and hidden patterns from educational datasets for decision-making and student performance analysis [1]. With the advancement of machine learning and data min-ing techniques, researchers have focused on predicting student dropout at an early stage so that appropriate intervention strategies can be implemented to improve student retention and aca-demic success. Various machine learning algorithms have been applied in previous studies to identify students who are at risk of dropping out based on demographic, academic, behavioral, and socio-economic factors. Yuksel Turk et al. applied data mining techniques including Decision Tree, Neural Network, Naive Bayes, and k-Nearest Neighbor (k-NN) classifiers for predicting dropout students in an online education program and achieved significant prediction accuracy [2]. Sani et al. focused on dropout prediction among B40 students in higher education using machine learning approaches and highlighted the importance of socio-economic and academic attributes in dropout prediction [3]. Formia et al. utilized data mining methods to characterize university dropout patterns and identify the major factors affecting student attrition [4]. Similarly, Aulck et al. used predictive models for higher education dropout prediction and demonstrated that machine learning algorithms can effectively identify students at risk of leaving their studies [5]. Sallan and Behal proposed an Enhanced Machine Learning Algorithm (EMLA) for student dropout prediction, which im-proved the classification performance compared to traditional machine learning methods [6]. Chaplot et al. introduced the Student Attrition Predictor (SAP) framework for predicting student attrition using educational datasets and machine learning techniques [7]. Ensemble learning methods have also shown promising results in educational prediction systems. Lagman et al. improved classification accuracy for student graduation prediction by using ensemble models and combining multiple classifiers [8]. Ademi and Loshkovska proposed an early detection system for identifying dropouts in e-learning environments using data mining techniques [9]. Hassan and Mirza applied machine learning algorithms for school dropout prediction and emphasized the role of predictive analytics in educational systems [10]. Umer presented a dropout prediction model for MOOCs environments to identify students at high risk of discontinuing online courses [11]. Bayesian classifiers have also been extensively used for academic performance prediction, where Sundar conducted a comparative study using Bayesian Network classifiers for predicting students' academic performance [12]. Wan Yaacob et al. employed data mining techniques for predicting student dropout in higher institutions and achieved effective classification performance using rates, academic failure, and student attrition. Researchers have applied various machine learning and data mining techniques to identify students at risk of dropping out and to improve educational decision-making processes [2–15]. These studies mainly focus on academic, demographic, behavioral, and socio-economic factors that influence student retention and academic success. Yuksel Turk et al. [2] explored dropout prediction in an online Information Technologies Certificate Program using data mining techniques. The study utilized variables such as gender, age, educational level, online learning readiness, self-efficacy, prior online experience, and occupation for dropout classification. Four machine learning algorithms, including k-Nearest Neighbor (k-NN), Decision Tree (DT), Naive Bayes (NB), and Neural Network (NN), were applied using 10-fold cross-validation. The results showed that the k-NN classifier achieved the highest sensitivity score of 87also identified self-efficacy, readiness for online learning, and previous online experience as significant factors affecting student dropout prediction. Sani et al. [3] focused on dropout prediction among B40 students in Malaysian higher education institutions. The study highlighted socio-economic disparities among low-income students and developed Decision Tree, Random Forest, and Artificial Neural Network models for predicting student attrition. Comparative analysis revealed that the Random Forest algorithm achieved the highest performance with an accuracy of 95.93%, precision of 97.10%, recall of 81.26%, and F-measure of 88.50%. Their findings demonstrated the effectiveness of Random Forest in identifying students at risk of dropping out. Formia et al. [4] utilized educational datasets [13]. Monllao´ Olive´ et al. proposed a data mining techniques to analyze university dropout patterns

supervised learning framework that used assessment-based features to identify students at risk of dropping out in MOOCs environments [14]. Furthermore, Gil et al. developed a Neural Network-based approach for predicting early students with a high risk of university dropout and demonstrated the effectiveness of deep learning techniques in educational prediction systems [15]. Inspired by these previous studies, this research aims to perform a comparative analysis of various machine learning algorithms including Logistic Regression, Decision Tree, Naive Bayes, Neural Network, k-Nearest Neighbors, Bayesian Network, Support Vector Machine (SVM), Random Forest, WAODE, AODESR, and Ensemble of Multiple Learners Algorithm (EMLA) for student dropout prediction. The performance of these algorithms is evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Additionally, feature importance analysis is conducted to identify the most influential factors affecting student dropout prediction. The findings of this study can assist educational institutions and policymakers in implementing effective early intervention strategies to reduce dropout rates and improve student retention.

2. Related Work and Discussion

dropout prediction has become an important research area in Educational Data Mining (EDM) and machine learning because of the increasing concerns regarding low graduation and identify the factors responsible for student attrition. The study emphasized the importance of educational datasets in understanding dropout behavior and supporting institutions in improving retention strategies for students facing academic difficulties. Aulck et al. [5] investigated student dropout pre-diction in higher education using predictive machine learning models. Their study demonstrated that machine learning techniques can effectively identify students who are at risk of leaving their studies by analyzing educational and behavioral data. The findings highlighted the usefulness of predictive analytics in educational systems for improving student retention and academic planning. Sallan and Behal [6] proposed an Artificial Intelligence (AI)-based expert system for predicting college dropouts using data mining classification techniques. The study evaluated Decision Stump, NDTREE, and Enhanced Machine Learning Algorithm (EMLA) on a benchmark student dropout dataset. The results showed that the EMLA model improved prediction accuracy compared to traditional machine learning approaches. Their work also highlighted the practical applications of AI-based systems for educational institutions in improving educational quality and supporting at-risk students. Chaplot et al. [7] introduced the Student Attrition Predictor (SAP) framework for predicting student attrition in MOOCs

environments. The framework utilized Artificial Neural Net-works and sentiment analysis techniques to analyze students' online activities and emotional responses. The study also developed a web-based system that allows educators to identify students at risk of dropping out and implement timely intervention strategies to improve course completion rates. Lagman et al. [8] proposed ensemble learning methods for student graduation prediction. The study focused on improving classification accuracy by combining multiple machine learning models into a single ensemble framework. The findings demonstrated that ensemble models achieved better prediction performance and supported educational institutions in identifying students at risk of delayed graduation. Ademi and Loshkovska [9] developed an early dropout detection system for e-learning environments using Moodle logs and data mining techniques. The researchers applied Decision Tree and Bayesian Network algorithms to predict dropout cases based on students' online activities before the first examination. The study revealed that the Decision Tree algorithm achieved higher overall prediction accuracy, whereas the Bayesian Network classifier showed better specificity in identifying failing students. Hassan and Mirza [10] investigated school dropout prediction in develop-ing countries using supervised machine learning algorithms. The study analyzed socio-economic and school-related factors affecting dropout rates and presented a comparative analysis of five classification algorithms for identifying students at risk of leaving school. The findings emphasized the role of predictive analytics in developing effective dropout prevention strategies. Umer [11] focused on dropout prediction in Massive Open Online Courses (MOOCs) using students' online activity data. The study employed machine learning algorithms such as Naive Bayes, Random Forest, Logistic Regression, and k-Nearest Neighbor for classification. Among these algorithms, Logistic Regression achieved the highest prediction accuracy, demonstrating its effectiveness in predicting student attrition in online learning environments. Sundar [12] conducted a comparative study on Bayesian Network classifiers for predicting students' academic performance. The study focused on identifying unmotivated students before the final examination stage and generating predictive models for detecting potential dropouts and students requiring additional academic support. The research also highlighted the importance of performance prediction in scholarship selection and academic planning. Wan Yaacob et al. [13] applied data mining techniques for predicting student dropout in higher education institutions. The study demonstrated that machine learning algorithms can effectively analyze educational datasets and provide accurate classification results for identifying students at risk of dropping out. Monllao´ Olive´ et al. [14] proposed a supervised learning framework for identifying students at risk of dropping out in MOOCs environments. The framework utilized assessment-based features and supervised learning methods to improve dropout prediction accuracy and support early intervention strategies for online learners. Gil et al. [15] developed a Neural Network-based approach for predicting students with a high risk of university dropout. Their study demonstrated that deep learning techniques can effectively analyze educational data and improve the prediction of student attrition in higher education systems. Although previous studies have achieved significant progress in student dropout prediction using ma-chine learning and data mining techniques, several limitations still exist. Many studies focus only on specific educational environments such as MOOCs or online learning systems, while others rely on limited datasets or a small number of classification algorithms. Furthermore, some predictive models lack comparative analysis among multiple machine learning approaches. Therefore, there is a need for a comprehensive comparative framework that evaluates different machine learn-ing algorithms for student dropout prediction using standard evaluation metrics such as accuracy, precision, recall, and
F1-score to identify the most effective predictive model for educational systems.

3. Methodology

Fig 1 presents the overall methodology adopted for student dropout prediction using machine learning algorithms. The proposed framework consists of dataset collection, data preprocessing, loading certain algorithms, performance evaluation, and comparative analysis. Each stage of the methodology is described in detail below.

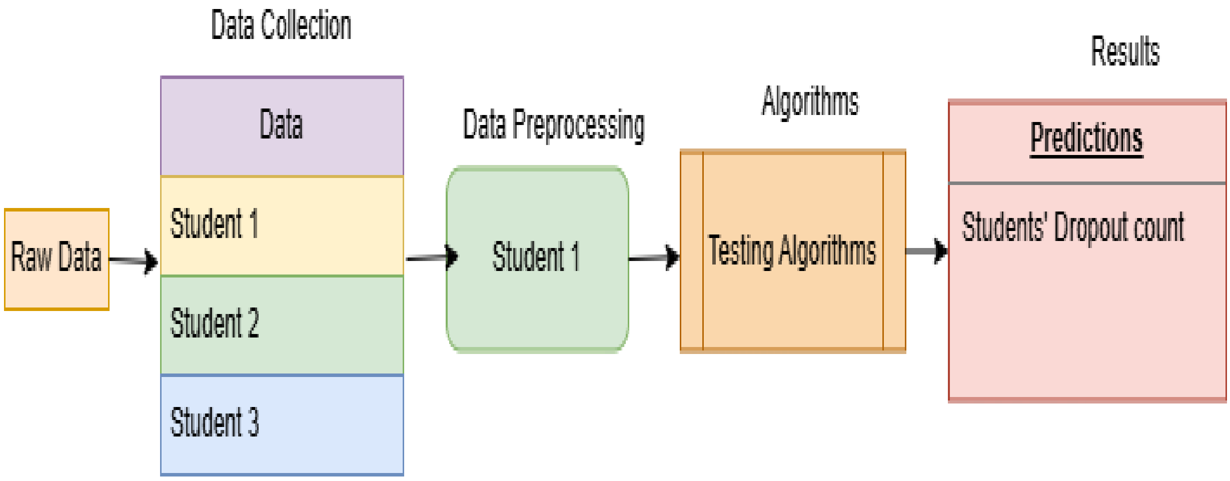

Fig. 1. Performance Comparison of Machine Learning Models

a. dataset

This study employs a publicly accessible educational dataset sourced from the UCI Machine Learning Repository, a reputable platform commonly utilized by researchers for evaluating machine learning models. The dataset, named "Predict Students Dropout and Academic Success," includes extensive information about students, such as demographic characteristics, academic achievements, attendance data, assignment outcomes, socio-economic status, and online learning behaviors. These characteristics are crucial for comprehending student behavior and recognizing factors linked to the risk of dropping out and academic achievement. Utilizing a publicly accessible dataset improves the trustworthiness, clarity, and replicability of the suggested research framework. The dataset features records of both dropout and non-dropout students, allowing machine learning algorithms to accurately identify classification patterns and differentiate between students at risk of discontinuing their studies and those likely to finish their education successfully. Despite the dataset featuring many educational attributes, this research primarily emphasizes key factors like academic performance, attendance rates, assignment grades, financial situation, engagement levels, and prior academic records, since these variables are closely associated with student retention and dropout forecasts in higher education systems. To guarantee precise experimental assessment, the dataset was segmented into training and testing subsets for model development and evaluation. Moreover, suitable preprocessing methods such as data cleansing, normalization, feature selection, and addressing imbalanced class distributions were utilized to enhance the predictive accuracy, robustness, and dependability of the machine learning models. Employing this dataset establishes a solid basis for creating intelligent predictive systems that can aid educational institutions in the early detection of at-risk students and the execution of prompt intervention strategies.

b. Dataset Preprocessing

After data collection, preprocessing techniques are applied to improve the quality and consistency of the dataset. Initially, missing values are handled using appropriate data cleaning methods, while duplicate and inconsistent records are removed. After that nominal values are converted into binary values. Outlier detection techniques are also applied to reduce noise and improve model reliability. Furthermore, categorical attributes are transformed into machine-readable numerical representations using encoding methods, whereas numerical attributes are normalized to maintain consistency among feature scales. Feature engineering techniques are also performed to identify the most relevant attributes and generate additional features that can enhance the predictive capability of machine learning models. This preprocessing stage ensures that the dataset is suitable for efficient model training and accurate prediction results.

c. Machine Learning Models

Several machine learning algorithms were implemented for comparative analysis of student dropout prediction. These include Logistic Regression, Decision Tree, Random Forest, Multilayer Perceptron (MLP), k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM). Each model was trained using the processed educational dataset to identify students who are at risk of dropping out and to evaluate the effectiveness of different classification techniques for educational data mining. Logistic Regression was used as a baseline classification model because of its simplicity, interpretability, and effectiveness in classification problems. Decision Tree and Random Forest were implemented due to their capability to handle complex relationships among student-related attributes and generate efficient classification rules. Multilayer Perceptron (MLP), a neural network-based approach, was applied to capture nonlinear patterns and hidden relationships within the educational dataset. Similarly, k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) were employed to compare their predictive performance and classification capabilities for student dropout prediction.

d. Experimental Design and Implementation

For experimental analysis, the publicly available UCI educational dataset was divided into training and testing subsets to evaluate the predictive performance of the selected machine learning algorithms. To ensure robustness, reliability, and better generalization capability of the models, the 10-Fold Cross-Validation technique was employed during the training and evaluation process. In addition, hyperparameter tuning was performed to optimize classifier performance, minimize overfitting, and improve overall prediction accuracy. Several machine learning models including Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, Multilayer Perceptron (MLP), and k-Nearest Neighbors (KNN) were experimentally evaluated for student dropout prediction. The effectiveness of each classifier was measured using standard evaluation metrics such as Accuracy, Precision, Recall, F1-Score, Sensitivity, Specificity, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and Area Under the Precision-Recall Curve (AUC-PR). These evaluation measures provide a comprehensive assessment of classification performance, particularly for educational datasets where class imbalance exists between dropout and non-dropout students. The experimental results demonstrated that the Support Vector Machine (SVM) achieved the highest classification performance with an accuracy of 93.80%, significantly out per-forming the remaining models. Comparative analysis further indicated that ensemble and advanced machine learning techniques provided better robustness and predictive capability for identifying at-risk students. The generated performance comparison graph visually illustrates the superiority of the SVM model over other classifiers, highlighting its effectiveness for educational data mining and student retention analysis.

e. Comparative Analysis and Prediction

After training and evaluation, a comparative analysis was performed to identify the best-performing machine learning algorithm for student dropout prediction. Feature importance analysis was also conducted to determine the most influential factors contributing to student attrition. The obtained results can help educational institutions and policymakers develop early warning systems capable of identifying at-risk students and implementing timely intervention strategies to improve student retention and academic success. The performance comparison of different classifiers is presented in Table I. Among all evaluated models following 10-Fold Cross Validation, the Support Vector Machine (SVM) achieved the highest performance, obtaining an accuracy of 93.80%, precision of 0.939, recall of 0.938, and F1-score of 0.938. In contrast, Logistic Regression showed the lowest performance with an accuracy of 36.40%. Other models including Random Forest, Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), and Decision Tree achieved comparatively lower accuracies around 37%, although some models demonstrated high recall values.

The experimental findings demonstrate that SVM significantly outperformed the remaining classifiers for student dropout prediction. Therefore, the proposed framework pro-vides an efficient and practical solution for educational data mining and predictive analytics in higher education systems.

$y = f(X) = argc0, 1maxP(cX)$ (1)

TABLE I: Performance Comparison of Machine Learning Classifiers

S.No	Classifier	Precision	Recall	F1-Score	Accuracy
1	SVM (10-Fold CV)	0.939	0.938	0.938	93.80%
2	Random Forest (10-Fold CV)	0.378	0.947	0.540	37.80%
3	Multilayer Perceptron (10-Fold CV)	0.378	0.861	0.525	37.80%
4	K-Nearest Neighbors (10-Fold CV)	0.383	0.910	0.539	37.80%
5	Decision Tree	0.376	1.000	0.546	37.60%
6	Logistic Regression (10-Fold CV)	—	0.364	—	36.40%

Equation 1 was used by the implemented machine learning models to classify students into dropout and non-dropout categories based on their educational attributes. The models analyzed input features such as academic performance, at-tendance, assignment scores, and socio-economic background, and predicted the class with the highest probability.

4. Experimentation, Results and Discussion

A comprehensive experimental procedure was designed to evaluate the effectiveness of the proposed machine learning framework for student dropout prediction. The collected educational dataset was divided into training and testing sets to assess the predictive capability of different machine learning algorithms. Since educational datasets may contain imbalanced class distributions, where dropout students are fewer com-pared to non-dropout students, suitable data preprocessing and balancing techniques were applied to improve classification performance and model reliability. Hyperparameter tuning was also performed to optimize the performance of the selected algorithms and minimize overfitting during the training process. Various machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were evaluated using standard performance metrics such as Accuracy, Precision, Recall, and F1-Score. The experimental results demonstrated differences in classification performance among the implemented models.

Fig. 2 presents the performance comparison of different machine learning models using classification accuracy. The bar chart clearly illustrates that the Support Vector Machine (SVM) significantly outperformed all other classifiers with an accuracy of 93.8%, indicating its strong predictive capability for student dropout prediction. In contrast, Random For-est, Multilayer Perceptron (MLP), and K-Nearest Neighbors (KNN) achieved similar accuracies of 37.8%, while Decision Tree and Logistic Regression achieved 37.6% and 36.4% accuracy, respectively. The graph visually highlights the substantial performance gap between SVM and the remaining models, demonstrating that SVM provided the most reliable and robust classification results for the given educational dataset.

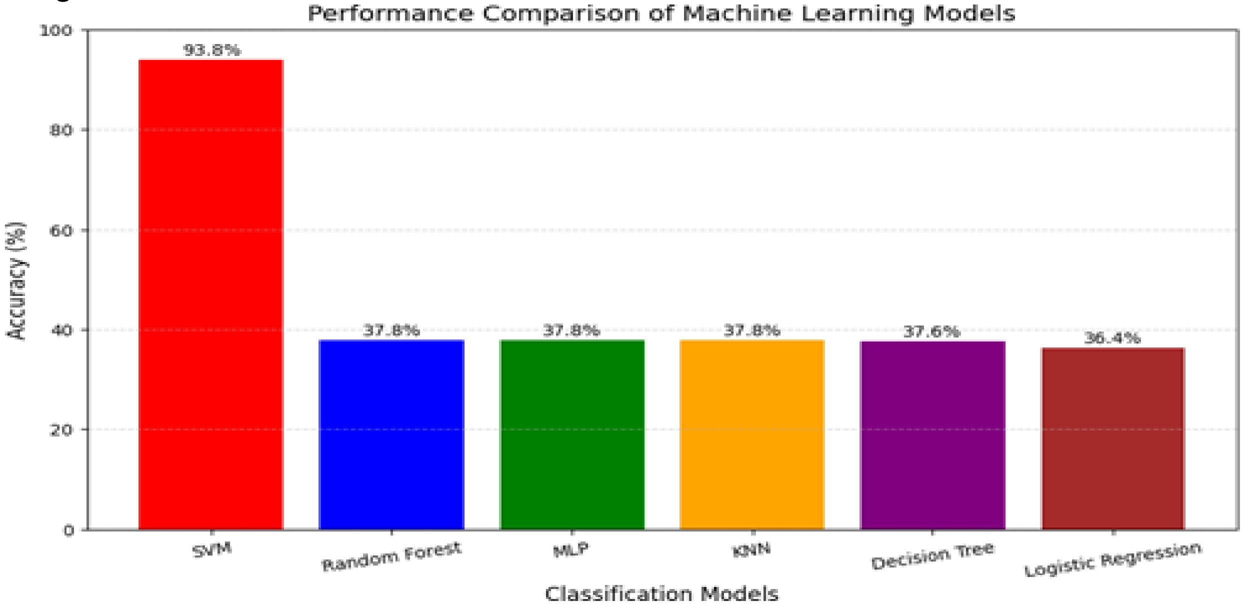

Fig. 2. Performance Comparison of Machine Learning Models

The findings further revealed that academic performance, attendance records, assignment results, engagement level, and socio-economic background were among the most significant factors affecting student dropout prediction. Comparative analysis indicated that the Support Vector Machine (SVM) outperformed the remaining classifiers in terms of robustness and predictive capability. The obtained results confirm that machine learning-based prediction systems can effectively sup-port educational institutions in developing early intervention strategies to reduce dropout rates and improve overall student retention and academic success.

5. Conclusion

This study presented a comparative machine learning frame-work for predicting student dropout using a publicly available educational dataset obtained from the UCI Machine Learning Repository. The research focused on identifying students who are at risk of leaving their educational programs by analyzing academic performance, attendance records, assignment results, engagement level, demographic information, and socio-economic factors. Several machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM) were implemented and evaluated using standard performance metrics such as Accu-racy, Precision, Recall, and F1-Score.

Experimental results demonstrated that the Support Vector Machine (SVM) significantly outperformed the remaining classifiers and achieved the highest prediction accuracy of 93.80%, with precision, recall, and F1-score values of 0.939, 0.938, and 0.938 respectively. In comparison, Random Forest, MLP, KNN, Decision Tree, and Logistic Regression produced comparatively lower classification accuracies. The findings further revealed that academic performance, attendance per-centage, assignment scores, engagement level, and socio-economic background are among the most influential factors affecting student dropout prediction.

The obtained results confirm that machine learning-based prediction systems can effectively support educational institutions in identifying at-risk students at an early stage and implementing timely intervention strategies to improve student retention and academic success. Despite achieving promising results, certain limitations remain, including class imbalance issues, dataset dependency, and limited availability of educational attributes. Future research may focus on integrating deep learning techniques, ensemble learning approaches, and larger educational datasets to further enhance prediction accuracy, robustness, and practical applicability of student dropout prediction systems in real-world educational environments.

References

A. C. Lagman, L. P. Alfonso, M. L. I. Goh, J. A. P. Lalata, J. P. H. Magcuyao, and H. N. Vicente, “Classification Algorithm Accuracy Im-provement for Student Graduation Prediction Using Ensemble Model,” *International Journal of Information and Education Technology*, vol. 10, no. 10, pp. 723–727, 2020.

B. Ugalde and R. Venkateswaran, “A Research Travelogue Towards Edu-cational Data Mining,” *International Journal of Computer Applications*, vol. 179, no. 42, pp. 39–48, 2018.

D. Monllao´ Olive´, D. Q. Huynh, M. Reynolds, M. Dougiamas, and

D. S. Chaplot, E. Rhim, and J. Kim, “SAP: Student Attrition Predictor,” 2015.

D. Wiese, “A Supervised Learning Framework: Using Assessment to Identify Students at Risk of Dropping Out of a MOOC,” *Journal of Computing in Higher Education*, vol. 32, no. 1, pp. 9–26, 2020.

E. Yukselturk, S. Ozekes, and Y. K. Tu¨rel, “Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program,” *European Journal of Open, Distance and E-Learning*, vol. 17, no. 1, pp. 118–133, 2014.

G. Sallan and S. Behal, “Prediction of Student Dropout Using Enhanced Machine Learning Algorithm,” *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3821–3826, 2020.

L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, “Predicting Stu-dent Dropout in Higher Education,” *arXiv preprint arXiv:1606.06364*, 2016.

M. Gil, N. Reyes, M. Jua´rez, E. Espitia, J. Mosqueda, and M. Soria, “Predicting Early Students with High Risk to Drop Out of University Using a Neural Network-Based Approach,” 2013.

M. M. Hassan and T. Mirza, “Prediction of School Dropouts with the Help of Machine Learning Algorithms,” 2020.

N. Ademi and S. Loshkovska, “Early Detection of Drop Outs in E-Learning Systems,” *Academic Perspective Procedia*, vol. 2, no. 3, pp. 1008–1015, 2019.

N. Mohamad, “Drop-Out Prediction in Higher Education Among B40 Students,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 550–559, 2020.

N. S. Sani, A. F. M. Nafuri, Z. A. Othman, M. Z. A. Nazri, and K.

P. V. P. Sundar, “A Comparative Study for Predicting Student’s Academic Performance Using Bayesian Network Classifiers,” *IOSR Journal of Engineering*, vol. 3, no. 2, pp. 37–42, 2013.

R. Umer, “Prediction of Students’ Dropout in MOOC Environment,” *International Journal of Knowledge Engineering*, pp. 43–47, 2017.

References

S. Formia, L. Lanzarini, and W. Hasperue´, “Characterization of Univer-sity Drop-Out at UNRN Using Data Mining: A Study Case,” 2016.

W. F. Wan Yaacob, N. Mohd Sobri, S. A. M. Nasir, N. D. Norshahidi, and W. Z. Wan Husin, “Predicting Student Drop-Out in Higher Insti-tution Using Data Mining Techniques,” *Journal of Physics: Conference Series*, 2020.