Enhancing Student Performance Prediction in e-Learning Ecosystems Using Machine Learning Techniques

1. Introduction

E-learning ecosystems have become an essential component of modern education, offering flexibility and accessibility to learners worldwide. However, predicting student performance remains a challenge due to various influencing factors such as engagement, interaction, and assessment scores. This study explores the effectiveness of machine learning (ML) techniques in forecasting learner success using the Open University Learning Analytics Dataset (OULAD). The research analyzes four ML algorithms—Random Forest, Logistic Regression, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA)—to determine their predictive capabilities.

2. Dataset and Methodology

The study employs the OULAD dataset, which includes student demographics, course participation, assessments, and interactions with the online learning platform. The methodology involves:

- Data preprocessing: Handling missing values, normalizing data, and feature engineering.
- Feature selection: Identifying key attributes influencing student success.
- Model training and evaluation: Implementing ML algorithms and assessing performance using accuracy, precision, recall, and F1-score.

3. Machine Learning Models

- **Random Forest**: An ensemble learning method that builds multiple decision trees to improve accuracy and reduce overfitting.
- **Logistic Regression**: A statistical model used for binary and multi-class classification.
- **Support Vector Machine (SVM)**: A supervised learning model that finds the optimal hyperplane for classification.
- **Linear Discriminant Analysis (LDA)**: A dimensionality reduction technique that classifies data by maximizing class separability.

4. Results and Discussion

The comparative analysis of the four models revealed that Random Forest achieved the highest accuracy at **91%**, outperforming the other models in classifying students into “Distinction,” “Pass,” and “Fail” categories. Logistic Regression and SVM demonstrated moderate accuracy, while LDA showed lower performance due to its linear assumptions. The study highlights the effectiveness of ensemble learning in improving prediction accuracy.

5. Conclusion and Future Work

The study confirms that machine learning can significantly enhance student performance prediction in e-learning environments. Future work may explore deep learning techniques, feature augmentation, and real-time analytics for more dynamic student performance tracking. The integration of AI-driven recommendation systems could further personalize learning experiences, improving student success rates.

6. References

Fatima Ezzahraa EL Habti, Mustafa Hiri, Mohamed Chrayah, Abdelhamid Bouzidi, and Noura Aknin. "Enhancing Student Performance Prediction in e-Learning Ecosystems Using Machine Learning Techniques." *International Journal of Information and Education Technology*, Volume 15, Number 2, February 2025.