Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Binary Classification of Academic Outcomes Using Ensemble Learning and Neural Networks: A Case Study on OULAD

Lili Dwi Yulianto[1], Satriawan Desmana[2], Sutikman[3], Winarsih[4]
[1,4]Sistem Informasi, Universitas Nasional, Jl. Sawo Manila, Pejaten, Ps. Minggu Jakarta, [3]Bisnis Digital, Universitas Nasional, Jl. Sawo Manila, Pejaten, Ps. Minggu Jakarta, [2]Rekayasa Kemanan Siber, Politeknik Negeri Cilacap, Jl. Dr. Soetomo No.1, Sidakaya, Cilacap, Jawa Tengah

Article Info
Keywords:
Educational Data Mining (EDM),
OULAD,
Feature Selection,
Dense Neural Networks (DNN),
Machine Learning

ABSTRACT
The importance of academic classification in online learning platforms is increasingly recognized as it helps in assessing student performance, early detection of issues, and identifying factors that influence academic success. This study uses the Open University Learning Analytics Dataset (OULAD) to predict students' academic success in various classification areas, including Distinction vs Non-Distinction, Withdrawn vs Non-Withdrawn, Pass vs Non-Pass, and Pass vs Fail. The aim of this research is to compare machine learning and deep learning techniques, such as Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier, with a deep learning model based on Dense Neural Networks (DNN) to produce the best possible predictions. Relevant features are also selected using feature selection and dimensionality reduction strategies, including autoencoders and Recursive Feature Elimination (RFE). The results show that LightGBM and Gradient Boosting perform best in several classifications, with an accuracy of 75.47% for Pass vs Fail. On the other hand, DNN requires further refinement but shows potential in handling more complex classifications. In addition to identifying students at risk of failing, this method provides a deeper understanding of the variables affecting academic success in online learning environments.

Corresponding Author:
Lili Dwi Yulianto
Sistem Informasi, Universitas Nasional
Jl. Sawo Manila, Pejaten, Ps. Minggu Jakarta
lilidwianto@gmail.com

INTRODUCTION

Educational Data Mining (EDM) aims to uncover hidden patterns in educational data that can enhance the quality of learning(Romero & Ventura, 2010). Educational institutions are increasingly using online learning systems, generating complex data about student activities, academic achievement, and involvement in both academic and extracurricular activities. The Open University Learning Analytics Dataset (OULAD) is one of the most widely used datasets in EDM to analyze student learning patterns(Alhothali et al., 2022). However, the main problem is the lack of early detection for students at risk of failing, leading to Non-Pass, Non-Distinction, or Non-Withdrawn statuses(Lemay et al., 2021). This makes it difficult to apply timely interventions. Additionally, the data available regarding students has not been maximally utilized, even though it could be used to identify at-risk students and improve the

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

accuracy of academic performance predictions. Therefore, this study aims to make effective use of this data and implement early detection using machine learning and deep learning to minimize the risk of academic failure(Al-Zawqari et al., 2022; Buenaño-Fernández et al., 2019).

The objective of this study is to classify students' academic outcomes into several categories, namely Distinction vs Non-Distinction, Withdrawn vs Non-Withdrawn, Pass vs Non-Pass, and Pass vs Fail. By analyzing students' behavioral patterns and activities during their learning period, this research aims to assist educational institutions in making better-informed decisions regarding additional support or guidance. This study employs machine learning and deep learning techniques to predict academic success, with a focus on early detection of students at risk of failing(Yağcı, 2022). To achieve this goal, the study utilizes important features through feature selection and dimensionality reduction methods, such as Recursive Feature Elimination (RFE), to enhance prediction accuracy(Jawad et al., 2022).

The feature engineering process, typically performed manually and often extending the study period, is a common challenge in predicting student academic achievement. Researchers can utilize machine learning algorithms like Random Forest and Artificial Neural Networks to predict academic outcomes using raw data, avoiding intricate feature engineering. These algorithms have shown impressive effectiveness, with accuracies ranging from 86% to 88%, in identifying students at risk. Data preparation or preprocessing is essential for traditional tabular data, significantly affecting model performance(Al-Zawqari et al., 2022; Yahya et al., 2021). Techniques such as Artificial Neural Networks (ANN) offer high classification accuracy, while algorithms like Random Forest, combined with statistical techniques, can enhance model quality. For instance, key variables from e-learning platforms can be automatically extracted using the Moodle plugin for model evaluation and training(Hasan et al., 2021; Trishna et al., 2019). The main challenges include ensuring the relevance and linearity of the data, verifying model performance, and applying the model to new data. Additionally, handling non-linear relationships between variables may require more sophisticated approaches, such as deep learning, to uncover hidden patterns(Renò et al., 2022).

Using the OULAD dataset, this study aims to predict student academic success in a number of classification areas, including Distinction vs. Non-Distinction, Withdrawn vs. Non-Withdrawn, Pass vs. Non-Pass, and Pass vs. Fail. This study uses a deep learning model based on Dense Neural Networks (DNN) in conjunction with machine learning and deep learning algorithms, such as Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier, to get the best possible predicted outcomes(Natras et al., 2022; Trishna et al., 2019). To improve model accuracy, methods like autoencoders and Recursive Feature Elimination (RFE) are also used to choose and lower the dimensionality of the most pertinent features. With this approach, the study not only aims to identify students at risk of failure but also seeks to provide a comprehensive analysis of the factors influencing students' academic performance on online learning platforms(Almulihi et al., 2022; Habibi et al., 2023; Zhang et al., 2024).

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

METHODS

Attempting to predict student academic outcomes in a variety of binary classification categories, including Pass vs. Non-Pass, Pass vs. Fail, Withdrawn vs. Non-Withdrawn, and Distinction vs. Non-Distinction, this work applies the OULAD dataset and a number of machine learning and deep learning techniques. This study uses a number of steps, including as data preparation, encoding, feature selection, and model evaluation, to get accurate results. The flowchart depicting the research process is displayed in the following figure:

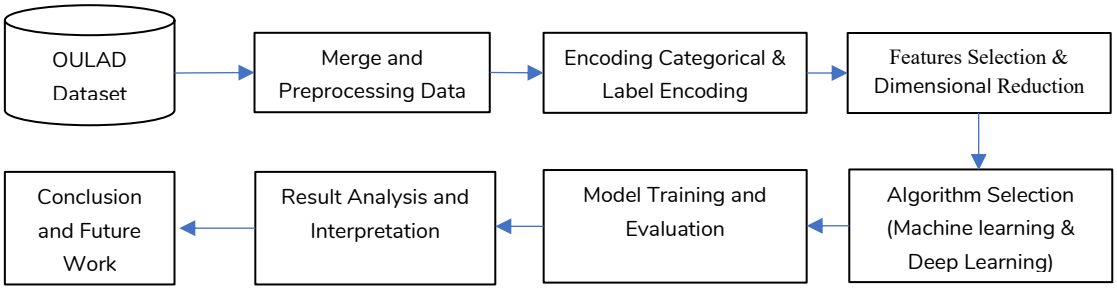

Figure 1. Research stage

This study begins with data collection and preprocessing from the OULAD dataset, including merging, handling missing values, and scaling. Categorical features are encoded using Category Encoding and Label Encoding. Feature selection is performed using Recursive Feature Elimination (RFE) and Dimensionality Reduction with Autoencoder. Various machine learning and deep learning models, including Random Forest, Gradient Boosting, AdaBoost, LightGBM, Voting Classifier, and Dense Neural Networks (DNN), are used for training and evaluation. Model performance is assessed using metrics like ROC-AUC, F1-score, recall, accuracy, and precision.

Dataset

The Open University Learning Analytics Dataset (OULAD) is a significant educational dataset, covering over 32,000 students from the Open University UK in 2013 and 2014. The dataset includes courses across disciplines like social sciences, engineering, science, technology, and mathematics (STEM). It provides demographic data, academic performance across different course sessions, and student interactions with the Virtual Learning Environment (VLE). Organized into seven CSV files, it includes information on courses, assessments, student demographics, and detailed VLE interactions, with over 10 million recorded interactions across 6,300 online resources. OULAD is commonly used to predict student failure or drop-out risks and assess student engagement with online learning materials, making it ideal for studying academic classification and prediction(Al-Zawqari et al., 2022; Jawad et al., 2022).

Data Categories.

Academic research often utilizes the Open University Learning Analytics Dataset (OULAD), an educational dataset. Data from more than 32,000 Open University UK students in 2013 and 2014 are included in this dataset. Some of the subjects covered by the courses in this collection are the social sciences and engineering, science, technology, and

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

mathematics (STEM). Data on student demographics, academic performance over multiple course sessions, and interactions with the Virtual Learning Environment (VLE) are all included in OULAD.

Data Features.

This research combines multiple datasets to create a new dataset consisting of 38 features and 1 target label for academic data analysis and prediction. The features include demographic data such as highest_education, imd_band, age_band, and gender (e.g., gender_F, gender_M). Engagement-related features include sum_click, After_Clicks, Before_Clicks, and date_registration. Additionally, features related to the student's place of residence and course modules, such as region_North Region and code_module_AAA, are included. The dataset also contains categorical features like Code_Category_STEM and Code_Category_Social_Science, which differentiate course categories.

The target label for classification is final_result_binary, which represents a binary classification of the students' final academic results. The four categories in this label are Pass, Fail, Withdrawn, and Distinction. The aim of this research is to predict academic outcomes in four binary classification scenarios: Pass vs Non-Pass, Pass vs Fail, Withdrawn vs Non-Withdrawn, and Distinction vs Non-Distinction. In the Pass vs Non-Pass scenario, students who pass or excel are grouped under "Pass," while those who fail or withdraw are categorized as "Non-Pass." The Pass vs Fail prediction compares graduates with students who fail, while the Withdrawn vs Non-Withdrawn prediction compares students who withdraw with those who remain enrolled. Finally, the Distinction vs Non-Distinction classification aims to predict which students will achieve high performance(Al-Zawqari et al., 2022).

Pre-Processing Data

The preprocessing stage begins with merging several subsets of the OULAD dataset to combine relevant information, including courses, assessments, studentInfo, and studentVle. This step is taken to ensure that all necessary student-related information can be used to create models using machine learning and deep learning algorithms. After the merger, null data handling was performed using the fillna and interpolate methods. The fillna method is used to fill empty values in features with default or relevant average values, while interpolate is used to estimate missing values based on data patterns around the empty points(Gnat, 2021). Next, in the Categorical Encoding process, features with specific categories such as gender and region are converted into numerical forms using One-Hot Encoding. Label Encoding is applied to the main binary label, final_result_binary, to convert categorical values into numbers, such as 0 and 1.

Features Importance

A machine learning feature selection consider named Recursive Feature Elimination (RFE) is used to train data with each attribute that is available, after which each feature is given a weight value. The least valuable feature will be eliminated. Until the ideal number of features the ones that most aid in the categorization process is attained, this process will be repeated. An autoencoder is a deep learning model that the reduces dimensionality by compressing input into a more accessible representation and then recreating the data. The

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

encoder, which converts the input data into a lower-dimensional representation, and the decoder, which tries to restore the data to its original form, are the two primary components of an autoencoder. An autoencoder makes it easier to train the model and avoid overfitting by lowering the dimensions of the data and capturing essential patterns without adding unnecessary information.

Algorithm Selection

In this study, several machine learning algorithms were employed to analyze the dataset and predict academic performance. The models used include Random Forest, which leverages ensemble learning to improve accuracy by aggregating multiple decision trees; Gradient Boosting, a powerful boosting method that builds models sequentially to correct errors made by previous trees; and AdaBoost, which enhances weak classifiers by focusing on the mistakes made on previous iterations. Additionally, the Light Gradient Boosting Machine (LightGBM) was utilized, known for its efficiency and speed in handling large datasets. The Voting Classifier was also applied to combine the predictions of multiple models for improved accuracy, while the Dense Neural Networks (DNN) were implemented to capture complex, non-linear relationships in the data through deep learning. These algorithms together form a robust approach to analyze and predict the academic outcomes of students.

RESULTS AND DISCUSSION

Initially, the research aimed to perform multiclass classification with four labels: Pass, Fail, Withdrawn, and Distinction. However, due to model complexity and subtle differences between categories, the results lacked sufficient accuracy. This led to a shift towards a binary classification approach with four scenarios: Pass vs Non-Pass, Pass vs Fail, Withdrawn vs Non-Withdrawn, and Distinction vs Non-Distinction. By simplifying the prediction task to two classes at each stage, this binary approach provided more stable and accurate results, especially in the presence of class imbalance or data variation.

Parameter for Features Importance

Table 1. Parameter for Autoencoder and RFE

method	configuration	description
Autoencoder	Encoding Dimensi: 5	Reduced dimension to decrease feature complexity.
	Activation: ReLU (encoder)	Uses ReLU activation in the encoder layer to model non-linear relationships.
	Output Activation: Sigmoid	Sigmoid activation in the output layer for reconstructing inputs on a scale of [0,1].
	Loss: Mean Squared Error	Mean Squared Error used as the loss function to measure reconstruction error.
	Optimizer: Adam	Adam optimizer used for stable and fast weight updates during training.
	Epochs: 50, Batch Size: 32	Training conducted over 50 epochs with a batch size of 32.

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

RFE	Estimator: Random Forest	Uses Random Forest as the estimator to determine feature importance.
	Number of Features: Top 5	Selects the top 5 most significant features that contribute to predictions.

With the setup at Table 1, the Autoencoder is guaranteed to decrease dimensionality while maintaining crucial feature information. The ReLU activation makes it possible to use non-linear modeling for efficient encoding and decoding, while the Adam optimizer stabilizes training. In contrast, RFE uses Random Forest to automatically identify the most important features, concentrating on the features that have the greatest impact on prediction accuracy. This combination increases model efficiency and reduces overfitting.

Pass vs Non-Pass.

Table 2 below presents the performance comparison results of five main algorithms (Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier) in the Pass vs Non-Pass binary classification scenario.

Table 2. Algorithm Comparison on Pass vs Non-Pass

		Random Forest	Gradient Boosting	AdaBoost	LightGBM	Voting Classifier
Accuracy	Standart	74,20%	75,50%	74,20%	75,50%	72,00%
	RFE	70,00%	72,00%	71,90%	71,90%	70,10%
	Autoencoder	68,30%	70,20%	69,60%	70,30%	68,80%
Precision	Standart	66,00%	64,90%	66,30%	64,90%	67,80%
	RFE	60,60%	62,40%	62,60%	62,00%	63,90%
	Autoencoder	59,10%	61,40%	60,60%	61,90%	60,20%
Recall	Standart	66,40%	76,60%	65,10%	75,50%	49,70%
	RFE	62,00%	69,80%	68,20%	70,60%	49,20%
	Autoencoder	60,20%	68,70%	67,20%	69,40%	48,80%
F1 Score	Standart	65,00%	69,60%	64,50%	69,40%	54,90%
	RFE	60,30%	64,20%	63,30%	64,60%	53,40%
	Autoencoder	58,90%	64,00%	62,40%	63,50%	53,20%
ROC AUC	Standart	82,80%	83,40%	82,80%	83,40%	82,40%
	RFE	78,40%	81,60%	81,20%	80,90%	79,30%
	AutoencodeR	77,80%	80,90%	80,50%	81,30%	79,10%

Table 2 above presents the performance of binary classification for Pass vs Non-Pass, evaluated using three approaches: Standard (without feature selection), RFE, and Autoencoder. In general, Gradient Boosting and LightGBM demonstrate strong performance,

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

particularly in Recall and ROC AUC metrics, indicating their effectiveness in detecting graduating students. However, the use of RFE and Autoencoder yields mixed results; while Autoencoder tends to improve Precision, both methods sometimes decrease accuracy, as seen with Random Forest, which dropped from 74.20% (Standard) to 68.30% (Autoencoder). Model testing was carried out using K-Fold Cross Validation with k = 5. The Voting Classifier showed overall lower performance, especially in Recall. These results suggest that Gradient Boosting and LightGBM are robust choices for predicting Pass vs Non-Pass, but feature selection and dimensionality reduction techniques should be applied cautiously to maintain optimal performance.

Pass vs Fail

Table 3 below presents the performance comparison results of five main algorithms (Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier) in the Pass vs Fail binary classification scenario.

Table 3. Algorithm Comparison on Pass vs Fail

		Random Forest	Gradient Boosting	AdaBoost	LightGBM	Voting Classifier
Accuracy	Standart	82,90%	83,81%	82,59%	83,26%	81,10%
	RFE	77,42%	79,10%	78,89%	78,98%	77,37%
	Autoencoder	63,60%	62,64%	59,89%	63,66%	63,51%
Precission	Standart	82,27%	80,91%	81,89%	80,95%	85,56%
	RFE	75,25%	76,10%	75,80%	75,99%	79,24%
	Autoencoder	64,47%	62,63%	60,69%	63,98%	63,57%
Recall	Standart	80,35%	85,20%	80,48%	83,89%	71,16%
	RFE	79,01%	83,03%	83,16%	82,78%	71,70%
	Autoencoder	51,75%	52,25%	56,54%	50,49%	52,30%
F1 Score	Standart	80,89%	82,83%	80,40%	82,11%	76,88%
	RFE F1	76,17%	78,47%	78,29%	78,32%	73,83%
	Autoencoder	50,73%	48,45%	50,90%	48,63%	49,62%
ROC AUC	Standart	90,91%	91,47%	91,13%	91,26%	90,16%
	RFE	87,08%	89,36%	88,99%	89,10%	87,30%
	Autoencoder	63,82%	65,38%	62,87%	66,71%	65,64%

The table above compares the performance of five classification models (Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier) for predicting Pass vs Fail, using K-Fold Cross Validation with k = 5. Overall, the Standard approach delivers the best performance across most metrics, with Gradient Boosting achieving the highest accuracy of 83.81% and a significant Recall value of 85.20%. LightGBM and AdaBoost also

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

show competitive results, particularly in Precision and ROC AUC. However, when the RFE (Recursive Feature Elimination) method is applied, model performance slightly decreases, with Gradient Boosting achieving the highest accuracy of 79.10%, while Recall remains high. In contrast, results from the Autoencoder show a significant performance drop across almost all metrics, with Random Forest and LightGBM experiencing accuracy decreases of up to 63.60% and 63.66%, respectively. Despite the poor performance of Autoencoder in the Pass vs Fail classification, Gradient Boosting consistently outperforms other models across all approaches.

Withdrawn vs Non-Withdrawn

Table 4 below presents the performance comparison results of five main algorithms (Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier) in the binary classification scenario of Withdrawn vs Non-Withdrawn.

Table 4. Algorithm Comparison Withdrawn vs Non-Withdrawn

		Random Forest	Gradient Boosting	AdaBoost	LightGBM	Voting Classifier
Accuracy	Standart	81,91%	82,04%	81,26%	81,73%	81,13%
	RFE	77,83%	79,50%	79,57%	79,22%	77,42%
	Autoencoder	63,90%	67,90%	68,21%	67,80%	67,41%
Precission	Standart	72,81%	74,13%	72,41%	71,92%	72,42%
	RFE	68,75%	72,79%	72,91%	72,00%	70,55%
	Autoencoder	40,35%	46,35%	34,65%	48,51%	49,46%
Recall	Standart	66,18%	64,31%	65,53%	67,45%	63,04%
	RFE	55,83%	56,99%	57,14%	57,25%	50,72%
	Autoencoder	31,80%	12,96%	16,69%	23,03%	15,73%
F1 Score	Standart	69,18%	68,70%	68,22%	69,46%	67,27%
	RFE	60,67%	62,83%	62,98%	62,64%	57,87%
	Autoencoder	34,08%	18,60%	18,33%	28,91%	18,27%
ROC AUC	Standart	87,65%	88,19%	87,83%	88,10%	87,15%
	RFE	83,20%	86,26%	86,12%	85,87%	83,75%
	Autoencoder	59,44%	62,55%	62,44%	63,38%	60,74%

Table 4 above shows the classification results of Withdrawn vs Non-Withdrawn using K-Fold Cross Validation with k = 5. In the Standard approach, Gradient Boosting and LightGBM showed the best performance, with Gradient Boosting achieving an accuracy of 82.04% and LightGBM excelling in Recall at 67.45%. When using RFE, performance slightly decreased, with the highest accuracy of 79.50% on Gradient Boosting, but Recall dropped in all models. The autoencoder showed a significant decline in all metrics, especially in Recall,

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

where Gradient Boosting only achieved 12.96%. Overall, Gradient Boosting and LightGBM remain consistently the best models across all approaches.

Distinction vs Non-Distinction

Table 5 below presents the performance comparison results of the five main algorithms (Random Forest, Gradient Boosting, AdaBoost, LightGBM, and Voting Classifier) in the Distinction vs Non-Distinction binary classification scenario.

Table 5. Algorithm Comparison Distinction vs Non-Distinction

		Random Forest	Gradient Boosting	AdaBoost	LightGBM	Voting Classifier
Accuracy	Standart	90,61%	90,61%	90,33%	90,50%	90,50%
	RFE	90,17%	90,63%	90,68%	90,57%	90,57%
	Autoencoder	89,93%	88,62%	90,72%	87,58%	90,72%
Precission	Standart	36,28%	43,32%	51,57%	33,57%	35,05%
	RFE	25,27%	33,58%	16,19%	29,04%	33,47%
	Autoencoder	23,74%	14,74%	0,00%	15,99%	0,00%
Recall	Standart	1,80%	1,77%	5,54%	2,50%	2,82%
	RFE	3,19%	1,14%	0,20%	1,30%	2,40%
	Autoencoder	3,26%	3,08%	0,00%	4,97%	0,00%
F1 Score	Standart	3,41%	3,36%	9,06%	4,64%	5,18%
	RFE	5,49%	2,19%	0,39%	2,45%	4,37%
	Autoencoder	5,01%	3,55%	0,00%	5,31%	0,00%
ROC AUC	Standart	79,29%	80,75%	80,67%	80,78%	80,02%
	RFE	73,50%	78,47%	78,25%	76,94%	75,21%
	Autoencoder	61,58%	62,62%	58,34%	60,20%	60,40%

Table 5 above presents the classification results for Distinction vs Non-Distinction using K-Fold Cross Validation with k = 5. In the Standard approach, all models achieved accuracy above 90%, with AdaBoost showing the highest Precision at 51.57%. However, when applying RFE, accuracy remained stable but there was a significant drop in both Precision and Recall, particularly in AdaBoost, which decreased drastically. The Autoencoder approach led to a more substantial decline, with models like AdaBoost and Voting Classifier showing Precision and Recall values of 0%, indicating their failure to predict the Distinction class. This suggests that the Autoencoder method is less effective for the Distinction vs Non-Distinction classification in this dataset.

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Architecture determination result.

Table 6. Parameter for Deep neural network

Layer (Type)	Output Shape	Activation	Parameters
Input Layer	(None, X_train_dl.shape[1])	-	0
Dense Layer	(None, 16)	ReLU	input_features * 16 + 16 (bias)
Dropout Layer	(None, 16)	-	0
Dense Layer (Output)	(None, 1)	Sigmoid	16 * 1 + 1 (bias)

The deep learning model architecture begins with an input layer that matches the number of features in the dataset. It is followed by a dense layer with 16 neurons and ReLU activation, along with L2 regularization to prevent overfitting. A Dropout layer with a 50% rate is applied to further regularize the model by randomly deactivating half of the neurons during training. The final output layer has a single neuron with a Sigmoid activation function, ideal for binary classification, producing an output between 0 and 1. The model is optimized using the Adam optimizer, with binary crossentropy as the loss function, and its performance is evaluated using accuracy, precision, and recall metrics. This architecture balances simplicity and regularization techniques to enhance generalization and minimize overfitting.

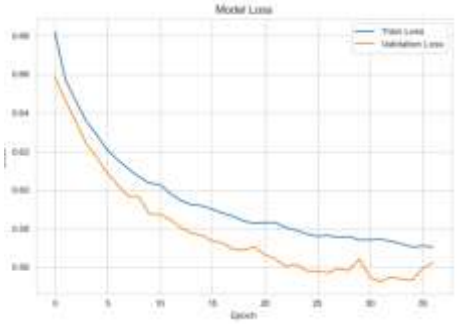

Figure 2. Model Loss Pass vs Non-Pass

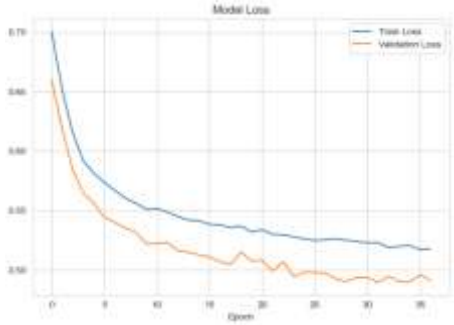

Figure 3. Model Loss Pass vs Fail

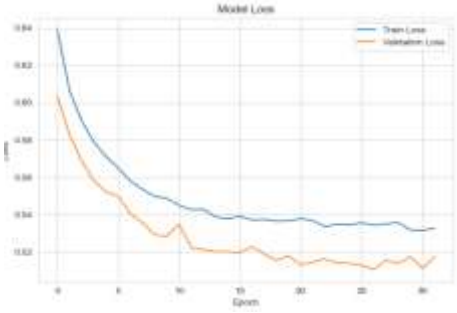

Figure 4. Model Loss Withdrawn vs Non-Withdrawn

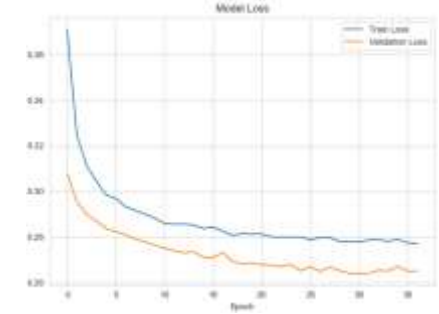

Figure 5. Model Loss Distinction vs Non-Distinction

In Figure 2, the Pass vs Non-Pass classification shows a decrease in both Train Loss and Validation Loss. Both losses decrease until around the 20th epoch, where Validation Loss

Jurnal Info Sains : Informatika dan Sains
Volume 15, Number 01, 2025, DOI 10.54209/infosains.v15i01
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

stabilizes, while Train Loss continues to decrease slightly. This decrease suggests that the model is improving its ability to minimize prediction errors. In Figure 3, both Train Loss and Validation Loss decrease rapidly at the start of training, stabilizing around the 10th epoch. Validation Loss remains lower than Train Loss throughout, indicating the model is adapting well to the training data while maintaining performance on validation data. In Figure 4, both losses decrease, with Validation Loss decreasing faster and staying lower than Train Loss throughout training. It reaches stability around the 10th epoch with slight fluctuations, showing that the model adapts well to data patterns without overfitting. In Figure 5, both losses decrease significantly in the early epochs and then stabilize at low values. Validation Loss stays slightly below Train Loss, suggesting good generalization without overfitting. However, the limited decrease after a few initial epochs indicates the model has reached its optimal point in classifying Distinction vs Non-Distinction.

Table 7. Result Confusion Matrix for Deep neural network

Classification Model	Accuracy	Precision	Recall	F1 Score
Pass vs Non-Pass	69,60%	62,09%	49,72%	55,22%
Pass vs Fail	74,70%	71,35%	76,79%	73,97%
Withdrawn vs Non-Withdrawn	74,40%	63,04%	37,25%	46,83%
Distinction vs Non-Distinction	90,90%	100,00%	0,00%	0,00%

From the confusion matrix results, the highest classification performance was observed in the Pass vs Fail category, with an accuracy of 74.70%, precision of 71.35%, recall of 76.79%, and an F1 Score of 73.97%. The high recall value indicates the model's ability to detect most "Fail" cases, while the balanced F1 Score reflects stable performance. In the Pass vs Non-Pass classification, the model achieved an accuracy of 69.60%, precision of 62.09%, and recall of 49.72%. The higher precision suggests the model is cautious in predicting "Pass," but struggles with detecting "Non-Pass" cases, indicating weaknesses in recognizing all data in this category.

For Withdrawn vs Non-Withdrawn, the model's performance decreased, with accuracy at 74.40%, precision at 63.04%, recall at 37.25%, and an F1 Score of 46.83%. The low recall highlights difficulty in detecting "Withdrawn" cases, suggesting the need for class balancing or additional tuning to improve performance. Finally, in Distinction vs Non-Distinction, despite an accuracy of 90.90% and 100% precision, the recall and F1 Score were 0.00%. This shows the model predicts predominantly "Non-Distinction" and fails to recognize "Distinction" cases. This indicates a class imbalance or the model's difficulty distinguishing the two classes, which may require balancing techniques or data augmentation to achieve optimal results.

CONCLUSION

This study highlights the role of Machine Learning (ML) and Deep Learning (DL) in Educational Data Mining (EDM) to predict student academic outcomes using the OULAD dataset. ML algorithms such as Gradient Boosting, LightGBM, and Random Forest performed well, with Gradient Boosting achieving the highest accuracy of 75.47% in the Pass vs Fail classification. LightGBM excelled in Pass vs Non-Pass classification, and Random Forest showed the best

performance in Withdrawn vs Non-Withdrawn. However, all algorithms struggled with the Distinction vs Non-Distinction classification, particularly in detecting high-performing students. Deep Neural Networks (DNN) showed competitive results but required adjustments like Dropout and Early Stopping to prevent overfitting. Although DNN holds potential, ML algorithms such as Gradient Boosting and LightGBM proved to be more efficient and stable for tabular data like OULAD, especially in handling class imbalance. The limitations of this research include the use of the OULAD dataset, which does not fully reflect the diversity and complexity of data from various educational institutions. Despite the use of several method combinations, challenges remain in classifying certain categories, such as Distinction vs Non-Distinction, which has a sparse data distribution. Recommendations for future research include developing models with more diverse data from various educational institutions and applying more sensitive algorithms such as ensemble or other deep learning techniques. Additionally, methods to handle imbalanced data can be explored further.

ACKNOWLEDGEMENT

We extend our deepest gratitude to those who have contributed to this research. We would like to extend our special thanks to Open University for providing the Open University Learning Analytics Dataset (OULAD), which has given us the opportunity to study academic data. We also appreciate the support from colleagues and research supervisors who provided guidance throughout this research.

REFERENCE

Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review. *Sustainability (Switzerland)*, *14*(10), 1–23. https://doi.org/10.3390/su14106199

Almulihi, A., Saleh, H., Hussien, A. M., Mostafa, S., El-Sappagh, S., Alnowaiser, K., Ali, A. A., & Refaat Hassan, M. (2022). Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction. *Diagnostics*, *12*(12), 1–17. https://doi.org/10.3390/diagnostics12123215

Al-Zawqari, A., Peumans, D., & Vandersteen, G. (2022). A flexible feature selection approach for predicting students' academic performance in online courses. *Computers and Education: Artificial Intelligence*, *3*(November), 100103. https://doi.org/10.1016/j.caeai.2022.100103

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability (Switzerland)*, *11*(10), 1–18. https://doi.org/10.3390/su11102833

Gnat, S. (2021). Impact of categorical variables encoding on property mass valuation. *Procedia Computer Science*, *192*, 3542–3550. https://doi.org/10.1016/j.procs.2021.09.127

Habibi, A., Delavar, M. R., Sadeghian, M. S., Nazari, B., & Pirasteh, S. (2023). A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for

flash flood susceptibility assessment. *International Journal of Applied Earth Observation and Geoinformation*, *122*(March), 103401. https://doi.org/10.1016/j.jag.2023.103401

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., & Sarker, K. U. (2021). Dataset of students' performance using student information system, moodle and the mobile application "edify." *Data*, *6*(11), 1–10. https://doi.org/10.3390/data6110110

Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability (Switzerland)*, *14*(22). https://doi.org/10.3390/su142214795

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, *2*(March), 100016. https://doi.org/10.1016/j.caeai.2021.100016

Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. *Remote Sensing*, *14*(15), 1–34. https://doi.org/10.3390/rs14153547

Renò, V., Stella, E., Patruno, C., Capurso, A., Dimauro, G., & Maglietta, R. (2022). Learning Analytics: Analysis of Methods for Online Assessment. *Applied Sciences (Switzerland)*, *12*(18), 1–10. https://doi.org/10.3390/app12189296

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Trishna, T. I., Emon, S. U., Ema, R. R., Sajal, G. I. H., Kundu, S., & Islam, T. (2019). Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier. *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, 1–7. https://doi.org/10.1109/ICCCNT45670.2019.8944455

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00192-z

Yahya, A. A., Sulaiman, A. A., Mashraqi, A. M., Zaidan, Z. M., & Halawani, H. T. (2021). Toward a better understanding of academic programs educational objectives: A data analytics-based approach. *Applied Sciences (Switzerland)*, *11*(20). https://doi.org/10.3390/app11209623

Zhang, P., Ma, Z., Ren, Z., Wang, H., Zhang, C., Wan, Q., & Sun, D. (2024). Design of an Automatic Classification System for Educational Reform Documents Based on Naive Bayes Algorithm. *Mathematics*, *12*(8), 1127. https://doi.org/10.3390/math12081127