



A Novel Stacking Framework for Student Success Prediction Using Combined Machine Learning and Deep Learning Approaches

Prepared By

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics
Midocean University
UAE

Supervisor Name:

Collage of Informatics, Midocean University, Comoros

Acknowledgements

Special thanks to

Abstract

The Educational field need a method to able to predict student performance to be able to take important actions in the right time to improve academic results of the students. In this thesis we present a stacking framework that merges machine learning (ML) and deep learning (DL) technologies to improve the accuracy of predicting student performance, by using a hybrid filter then wrapper feature selection technique that increase the speed of model training and improves understanding of features and predictive capability of the model. This framework makes it a good choice for tabular data like our dataset. The model architecture uses combination of machine learning algorithms as base learners as Linear Regression with $R^2 = 87.05\%$, Ridge Regression with $R^2 = 87.05\%$, Lasso Regression with $R^2 = 87.04\%$, Decision Tree with $R^2 = 87.14\%$, Random Forest with $R^2 = 87.22\%$, and K-Nearest Neighbors with $R^2 = 81.90\%$, and the meta-learner is Multi-Layer Perceptron (MLP). The stacking ensemble combines both the capabilities of the base models and the DL meta-learner to get a strong R-squared (R^2) score of 87.22%, which is better than the scores of individual baseline models. We incorporated data from Kaggle and preprocessing techniques, feature engineering, and model validation with standardized criteria. Then we used SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to make it easier to understand the stacking model predictions. Experimental findings show that the suggested stacking structure makes predictions. The framework supports educators and administrators in making better calls, with huge possibilities for customizing learning and preventing students from leaving school early.

Keywords: Student Prediction, Machine Learning, Deep Learning, Hybrid Feature Selection, Stacking Ensemble

Table of Contents

List of Abbreviations.....	viii
List of Figures.....	ix
List of Tables	x

Chapter 1: Introduction	1
1.1 Problem Definition.....	2
1.2 Thesis Objectives	3
1.3 Thesis Contribution.....	3
1.4 Thesis Organization	5
Chapter 2: Overview	6
2.1 Introduction.....	6
2.2 Artificial Intelligence (AI).	7
2.3 Machine Learning (ML).	7
2.4 Deep Learning (DL).....	8
2.5 Large Language Models (LLMs)	8
2.6 Explainable AI (XAI).....	9
2.7 Classification Techniques.....	9
2.7.1 Machine Learning Techniques.....	10
2.7.2 Deep Learning Techniques	14
2.8 Feature Selection	16
2.9 Evaluation Metrics	16
2.9.1 Mean Absolute Error (MAE)	17
2.9.2 Mean Squared Error (MSE).....	17
2.9.3 Root Mean Squared Error (RMSE).....	17
2.9.4 R-squared (R^2)	17

Chapter 3: Related Work.....	18
3.1 Introduction	18
3.2 History	18
Chapter 4: Proposed Approach	31
4.1 Introduction	31
4.2 Dataset Description.....	33
4.3 Preprocessing Dataset	34
4.4 Feature Selection Phase	35
4.4.1 Feature Selection Method	36
4.4.2 Wrapper-Based Performance Evaluation.....	37
4.5 Machine Learning Models	37
4.5.1 Traditional ML Models	38
4.5.2 Neural Network Models	40
4.6 Stacking Model Workflow	40
4.6.1 Base Model Prediction Synthesis	40
4.6.2 Meta-Learning Optimization	40
4.6.3 Final Prediction and Classification	41
4.7 Summary	41
Chapter 5: Experimental Results and Discussion	43
5.1. Introduction to the Chapter	43
5.2. Results of Individual Models.....	43
5.3. Results of Proposed Stacking Framework	44
5.4. Discussion of Findings	47
5.5. Limitations of the Proposed Work	47
5.6. Summary of the Chapter	47
Chapter 6: Conclusion and Future Work	49
6.1 Important Features	49
6.2 Future Work.....	50

References.....	51
------------------------	-----------

List of Abbreviations

- **Acc.:** Accuracy
- **ANN:** Artificial Neural Network
- **ML:** Machine Learning
- **DL:** Deep Learning
- **LLM:** Large Language Model
- **NLP:** Natural Language Processing
- **EDM:** Educational Data Mining
- **XAI:** Explainable Artificial Intelligence
- **SVM:** Support Vector Machine
- **RF:** Random Forest
- **DT:** Decision Tree
- **LSTM:** Long Short-Term Memory
- **CNN:** Convolutional Neural Network
- **RNN:** Recurrent Neural Network
- **LR:** Logistic Regression
- **KNN:** K-Nearest Neighbors
- **MAE:** Mean Absolute Error
- **MSE:** Mean Squared Error
- **R²:** R-Squared
- **SMOTE:** Synthetic Minority Oversampling Technique

List of Figures

Fig. 1 Artificial Intelligence (AI).....	7
Fig. 2 Machine Learning (ML).....	7
Fig. 3 Deep Learning (DL).....	8
Fig. 4 Large Language Models (LLMs).....	8
Fig. 5 Explainable AI (XAI).....	9
Fig. 6 Decision Tree (DT)	10
Fig. 7 Random Forest (RF).....	11
Fig. 8 K-Nearest Neighbors (KNN)	12
Fig. 9 Support Vector Classifier (SVC)	13
Fig. 10 Multi-Layer Perceptron (MLP)	14
Fig. 11 Recurrent Neural Network (RNN)	15
Fig. 12 Feature Selection	16
Fig. 13 Proposed Framework.....	32
Fig. 14 Results of Individual Models	44
Fig. 15 Highlight Improvements in Key Metrics	45

List of Tables

2.1	Student Performance Data Set Description
2.2	Students' Academic Performance Dataset Description

Chapter 1 Introduction

Introduction

In the last few years, evaluating and predicting student performance has become a very important task in the modern educational system, as having an accurate forecasting of student outcomes and performance helps educational institutions to take early interventions if needed, reduce dropout rates, and also to optimize and personalize the learning process for each student and as more students rely on Learning Management Systems (LMSs), every click leaves behind a trail of data whether it's opening a lecture, turning in homework, joining a forum discussion, or simply moving through the course which provides valuable insights into student behaviors and how it's effect their learning outcomes (Hernández-García et al., 2024) [1]. Also, the artificial Intelligence (AI) especially Machine Learning (ML) and Deep Learning (DL) has proven that they are highly effective in education as they can uncovering patterns within student data that might otherwise go unnoticed, by learning from students' behaviors and past performance, these models can capture complex relationships and accurately predict academic performance. (Guo, 2015) [2]. As we talk about DL, we find that Artificial neural networks (ANN) unique algorithm as it has the ability to capture non-linear and complex relationships, which makes them efficient when dealing with predicting academic outcomes with high accuracy for example Isljamovic and Suknovic (2014) found that ANNs could predict student performance with great accuracy using features like previous grades, attendance, and participation [3], also the integration of Large Language Models (LLMs), such as OpenAI's GPT and Google's BERT, has further advanced educational analytics by facilitating the analysis of unstructured data like student essays, forum discussions, and peer feedback (Alnasyan, 2024) [4]. And to enhance the model more, we find that feature selection plays an important role in developing

a robust model able to predict student performance by identifying the most informative features from our diverse student datasets, which mostly include many aspects like study habits, socioeconomic conditions, emotional factors, academic results, and more. That's why feature selection can increase model understandability, reduce overfitting, and training time. (Alshabandar et al., 2022) [5]. The educational field is very sensitive and needs great management, which is why wrong actions can cost a lot in terms of resources and the student learning process and their outcomes that's why explainable Artificial Intelligence (XAI) has become increasingly vital in educational contexts to ensure transparency and trust, that's why XAI Techniques like SHAP and LIME can make understanding the predictive models and their outputs much easier by showing educators which factors affect the student performance and help them take informed actions (Abid et al., 2025) [6]. By combining both ML, DL, and feature selection in this study our aim is to develop advanced and explainable models for classifying student performance. The tool we develop can help in shifting educational strategies from being reactive to proactive, and to empower education institutions to make informed decisions that support student achievement while advancing equity in learning.

1.1 Problem Definition

Despite the increasing availability of student data in academic institutions, many educational systems still struggle to identify students at risk of underperformance early enough to intervene effectively. Traditional evaluation methods often rely on post-exam results, which limit opportunities for timely support and personalized learning adjustments. Moreover, the complexity and diversity of educational data, including demographic, behavioral, attendance, and psychometric variables makes it challenging to isolate the most influential factors in predicting academic outcomes.

Including irrelevant or redundant features in predictive models can lead to overfitting, increased computational cost, and reduced interpretability. This highlights the need for efficient feature selection techniques that prioritize relevant attributes and support meaningful analysis. In parallel, classification algorithms based on machine learning (ML), and deep learning (DL) offer powerful tools for forecasting student success, yet their effectiveness depends heavily on data quality and model transparency.

Therefore, this research addresses the pressing need to develop a data-driven classification framework that combines advanced prediction models, enabling institutions to make timely, evidence-based decisions and understand the underlying drivers of student achievement.

1.2 Thesis Objectives

Using artificial intelligence techniques, the core objective is to design and evaluate intelligent systems that leverage machine learning (ML), and deep learning (DL) models arranged in a stacking ensemble to classify students based on diverse academic, demographic, behavioral, and psychometric data. By integrating feature-selection methods, the study aims to isolate the most influential attributes driving academic success, thereby boosting accuracy, lowering computational cost, and minimizing overfitting. The stacking framework will combine the complementary strengths of multiple base learners and a meta-learner to deliver robust, reliable predictions. Ultimately, the goal is to support data-driven interventions, enable early identification of at-risk students, and contribute to fair and effective academic planning across educational institutions.

1.3 Thesis Contribution

This thesis has the following main contributions:

1. Proposed a hybrid feature selection technique that employs both filter methods (such Pearson correlation, Chi-square, and Mutual Information) and wrapper methods (like recursive feature removal) to uncover things that might help predict student achievement. The approach reduces down on noise and extra information while keeping a good balance between speed, ease of understanding, and predictive capability for educational datasets.
2. Using ML basic learners including Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, and K-Nearest Neighbors to find patterns in student data. Uses a number of machine learning methods, including linearity in regression models and non-linearity in tree-based models, to make predictions more reliable and applicable to a wider range of situations.
3. A Multi-Layer Perceptron (MLP) meta-learner is used in the stacking ensemble to combine the predictions of the basis learners to capture complex non-linear relationships, it makes predictions more accurate than single models.
4. The suggested stacking ensemble that included ML base learners and MLP meta-learners performed better ($MSE = 17.3085$, $RMSE = 4.1604$, $MAE = 3.2145$, $R^2 = 0.8722$). By combining diverse model predictions better than baseline models, it makes it easier to estimate how well students will do.
5. Using SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to make the stacking framework's predictions easier to understand. This makes the model's decisions more trustworthy by

showing educators the most important factors (like academic performance and engagement metrics) that affect predictions, so they can take action.

6. The proposed stacking framework's exceptional interpretability and accuracy facilitates the precise identification of at-risk students, thereby providing educational institutions with actionable, data-driven solutions. This is in support of applications such as personalized learning and attrition prevention.

1.4 Thesis Organization

The remainder of the thesis is structured as follows:

Chapter 2 This chapter introduces core AI concepts such as ML, DL, LLMs, and XAI, and explains their roles in education and learning analytics. It then presents the foundations of Educational Data Mining (EDM), classification techniques, and types of learning algorithms. The importance of feature selection in predictive modeling is discussed. Finally, it highlights modern trends in AI-driven student analytics and their impact on proactive education.

Chapter 3 Examines recent academic research and comparative studies related to student performance prediction, machine learning applications in education, and the use of deep learning and language models. It identifies current research gaps that this thesis aims to address.

Chapter 4 Describes the research design and methodology, including data collection, preprocessing, algorithm selection, and the integration of feature selection techniques. It explains how different classification models and feature selection methods are applied and evaluated.

Chapter 5 Reports the experimental setup and presents results obtained from various classification models. The chapter includes a detailed performance analysis using accuracy, precision, recall, and F1-score. It also discusses the impact of feature selection on the model's performance.

Chapter 6 Summarizes the key findings, contributions, and implications of the study. It provides practical recommendations for educators and policymakers and outlines potential directions for future research in predictive analytics and educational AI systems.

Chapter 2 Overview

2.1 Introduction

The Artificial Intelligence (AI) and its subfields that contribute to Educational Data Mining (EDM). it consists of key concepts that it is relevant to the education field, namely Machine Learning (ML), Deep Learning (DL), Large Language Models (LLMs), and Explainable AI (XAI). Each concept is illustrated to show how it supports modern learning analytics. Also the machine learning can enable us to build predictive models to predict student performance, and with deep learning, can process complex educational data for deeper insights. Also, by handling natural language effectively, the LLMs can help in developing of educational tools .Also, XAI provides interpretable models to ensure transparency in learning analytics decisions.

In this chapter, we will focus on the key principles of the EDM and how the data-driven insights can be applied in the education field, also we will discuss regression techniques and key learning algorithms, as well as the importance of feature selection, so we can build accurate and robust predictive models. The chapter concludes by highlighting recent trends in AI-based student analytics, demonstrating how the integration of advanced AI techniques enables a transition from traditional reactive educational practices to more proactive, personalized, and insight-driven approaches.

2.2 Artificial Intelligence (AI)

Artificial Intelligence (AI) is a big field that tries to copy human abilities. In education, AI is used to make systems that can do things like grading, tutoring, or personalizing learning without a teacher always being there. Also AI have many subfields like machine learning (ML), deep learning (DL), and large language models (LLMs).

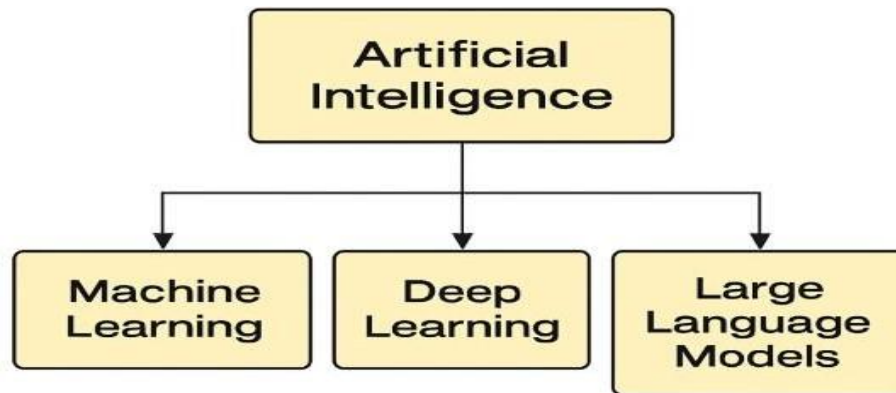


Fig. 1. Artificial Intelligence (AI)

2.3 Machine Learning (ML)

The machine learning (ML) is a part of AI that makes systems learn from data to get better and capture patterns from this data without direct programming. ML mostly works with tabular datasets. In education, ML is used to predict how students will perform, group them by achievement, and suggest useful learning materials [7].

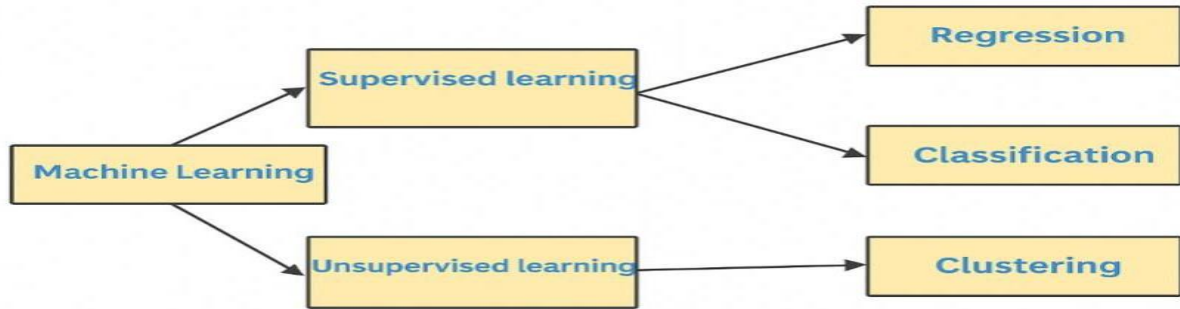


Fig. 2. Machine Learning (ML)

2.4 Deep Learning (DL)

The deep learning (DL) is considered a branch of ML that is based on neural networks with multiple hidden layers to form deep layers that can capture complex patterns in the data. In educational field, there are many applications of DL for example tasks like handwriting recognition, speech-to-text processing, and analyzing visual learning behaviors [8].

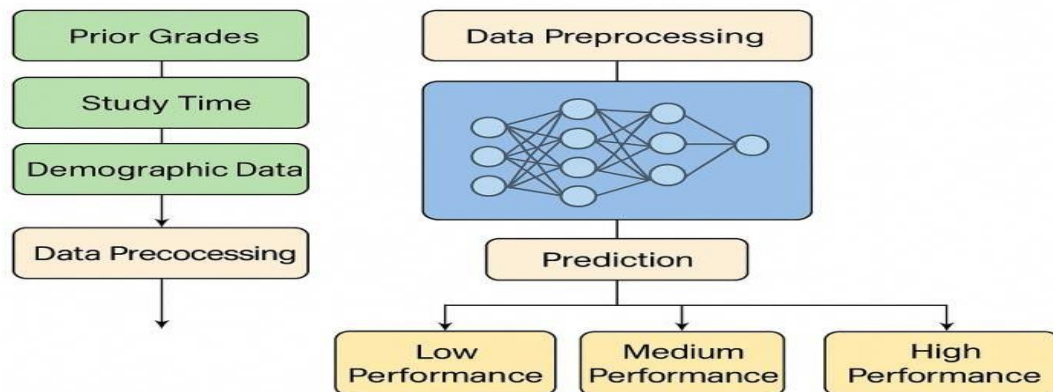


Fig. 3. Deep Learning (DL)

2.5 Large Language Models (LLMs)

Large Language Models (LLMs) like OpenAI's GPT and Google's BERT are advanced AI systems trained on vast amounts of text data. They are capable of understanding and generating human-like text. In education, LLMs are used to analyze student writing, conduct sentiment analysis, and generate automated feedback or tutoring responses [9].

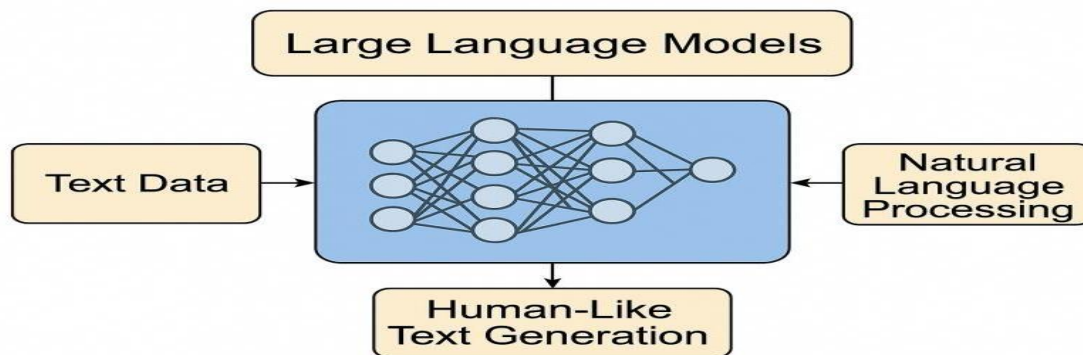


Fig. 4. Large Language Models (LLMs)

2.6 Explainable AI (XAI)

Explainable AI (XAI) refers to AI systems that offer transparent decision-making processes. XAI is crucial in education to ensure that predictions made by ML and DL models can be understood and trusted by educators and students. It helps in identifying which features influenced a prediction and supports ethical use of AI [10].

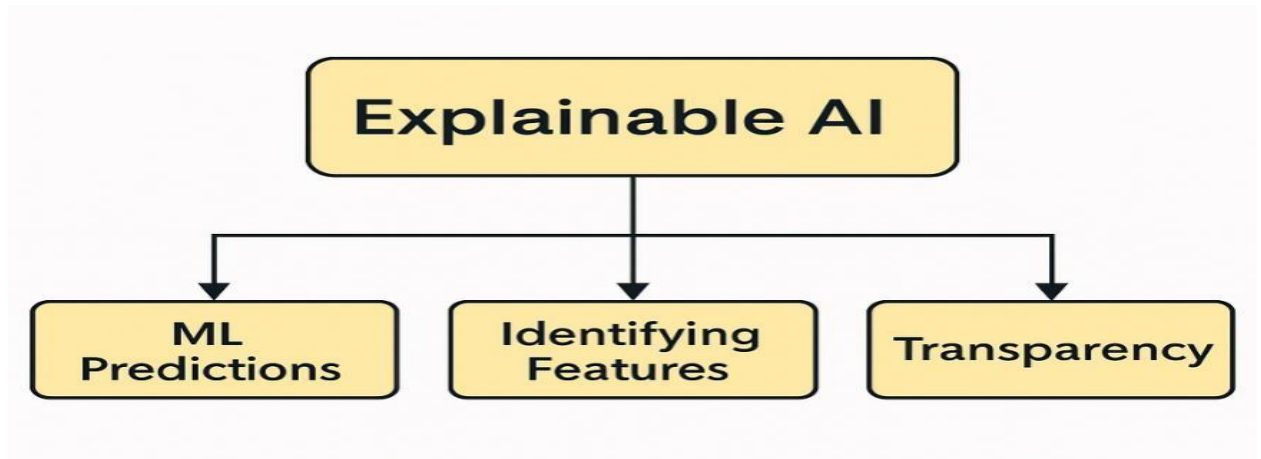


Fig. 5. Explainable AI (XAI)

2.7 Classification Techniques

Classification is a supervised learning technique used in EDM to predict student outcomes, such as pass/fail or grade categories. Common algorithms include Decision Trees, Support Vector Machines (SVM), Random Forests, and Neural Networks. These models can learn from historical data to classify new instances based on input features [11].

2.7.1 Machine Learning Techniques

Decision Tree (DT): A tree-structured classifier that splits the feature space using hierarchical decision rules, where each internal node represents a feature test, each branch represents an outcome, and each leaf node represents a class prediction. Decision trees are highly interpretable and can handle both numerical and categorical data without requiring feature scaling. However, they are prone to overfitting, especially with deep trees, and can be unstable as small changes in data may result in significantly different tree structures.

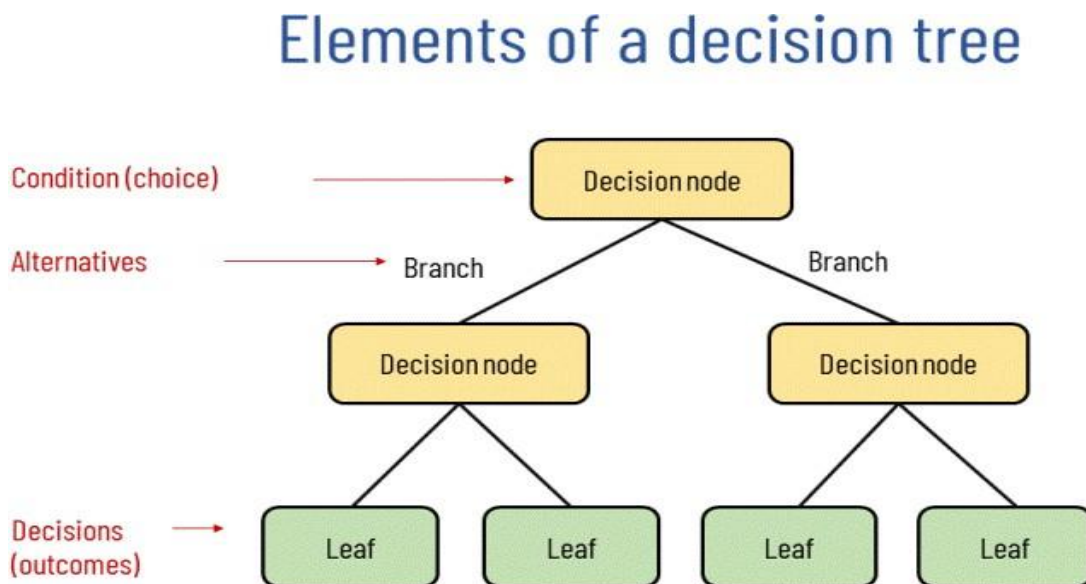


Fig. 6. Decision Tree

Random Forest (RF): An ensemble method that combines multiple decision trees through bootstrap aggregating (bagging), where each tree is trained on a random subset of the training data and features. The final prediction is determined by majority voting (classification) or averaging (regression) across all trees. Random forests reduce overfitting compared to individual decision trees, provide built-in feature importance measures, and are robust to outliers and missing values.

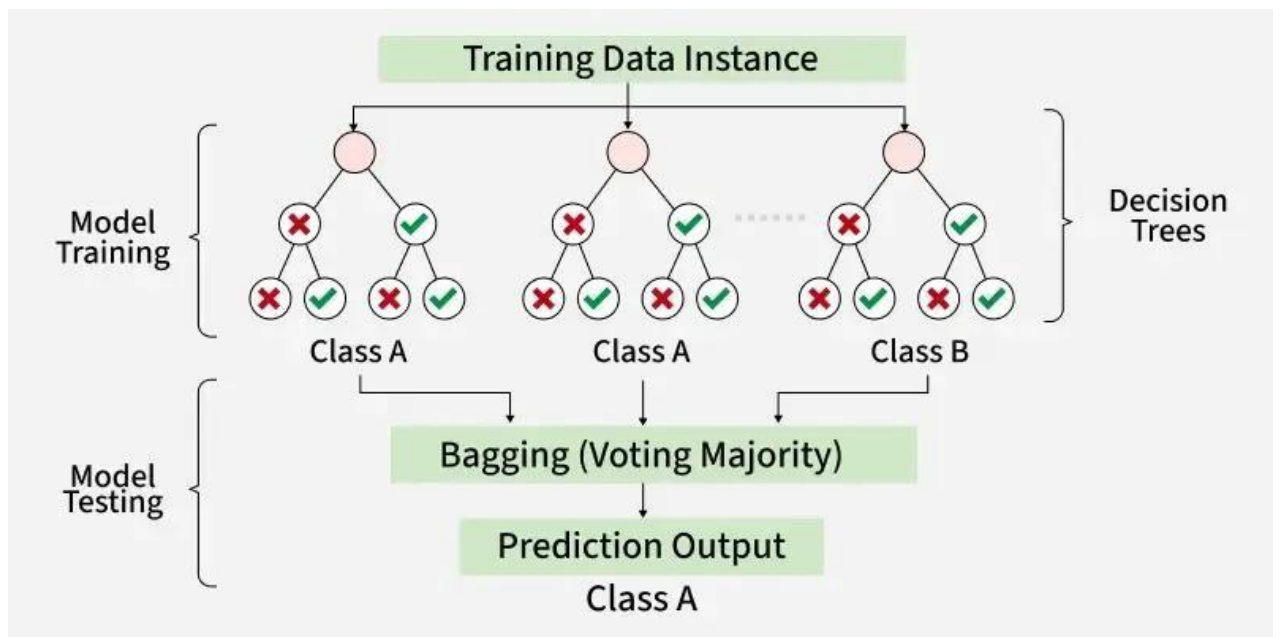


Fig. 7. Random Forest

K-Nearest Neighbors (KNN): A lazy learning algorithm that classifies instances based on the majority class among the k -nearest training samples in the feature space. KNN makes no assumptions about the underlying data distribution and can capture complex decision boundaries. However, it is computationally expensive during prediction, sensitive to the choice of k and distance metric, and performs poorly in high-dimensional spaces due to the curse of dimensionality.

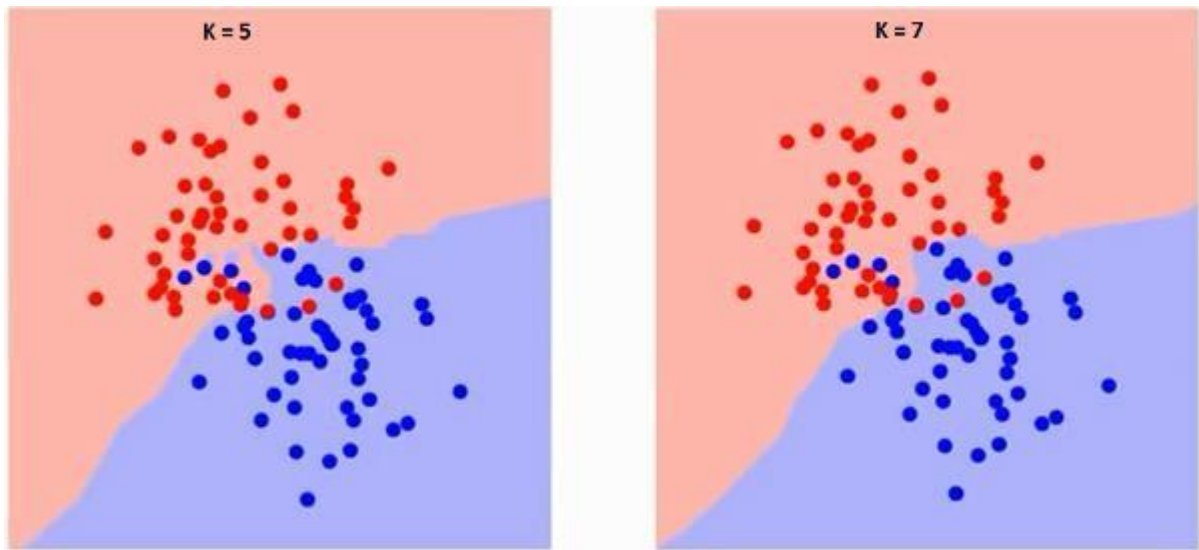


Fig. 8. KNN

Support Vector Classifier (SVC): A discriminative classifier that finds the optimal hyperplane to separate classes by maximizing the margin between different class boundaries. SVC can handle non-linearly separable data through kernel functions (polynomial, RBF, sigmoid) that map features into higher dimensional spaces. It is effective in high-dimensional spaces and memory efficiency but requires careful hyperparameter tuning and feature scaling.

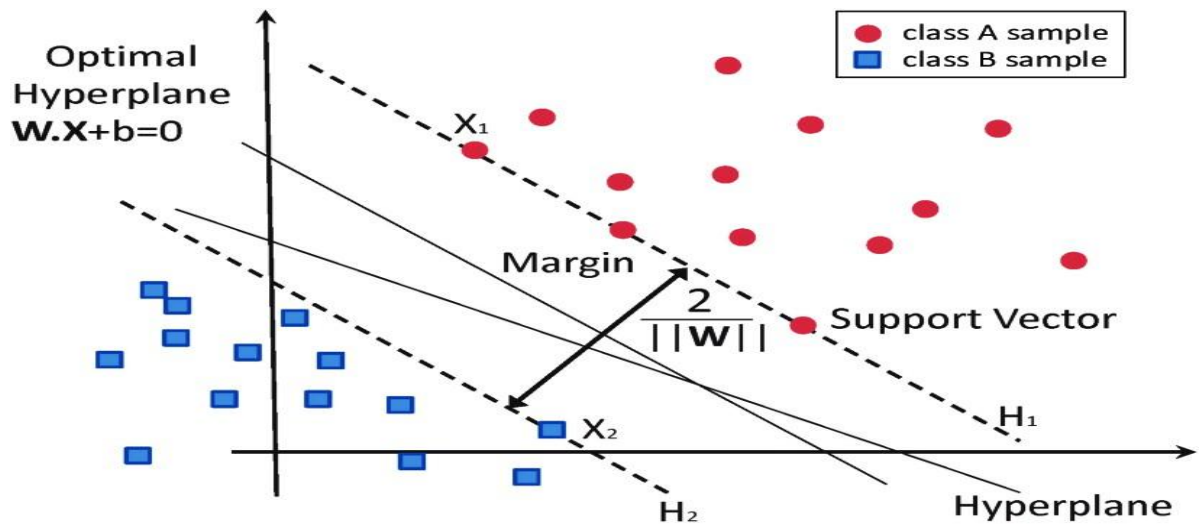


Fig. 9. SVC Algorithm

2.7.2 Deep Learning Techniques

Multi-Layer Perceptron (MLP)

It's a feed-forward artificial neural network (ANN) that is constructed from an input and output layer, and between them are multiple hidden layers. Each layer has a number of neurons that are connected to the next layer. Also, each neuron has its own weight and bias, which are considered the learnt parameters of the algorithm. Also, between each layer, there is an activation function, which makes the MLP able to learn from complex nonlinear relationships through backpropagation. Also, to enhance the model performance and reduce the overfitting, we need a careful architecture design and a regularization technique to reach the optimal performance.

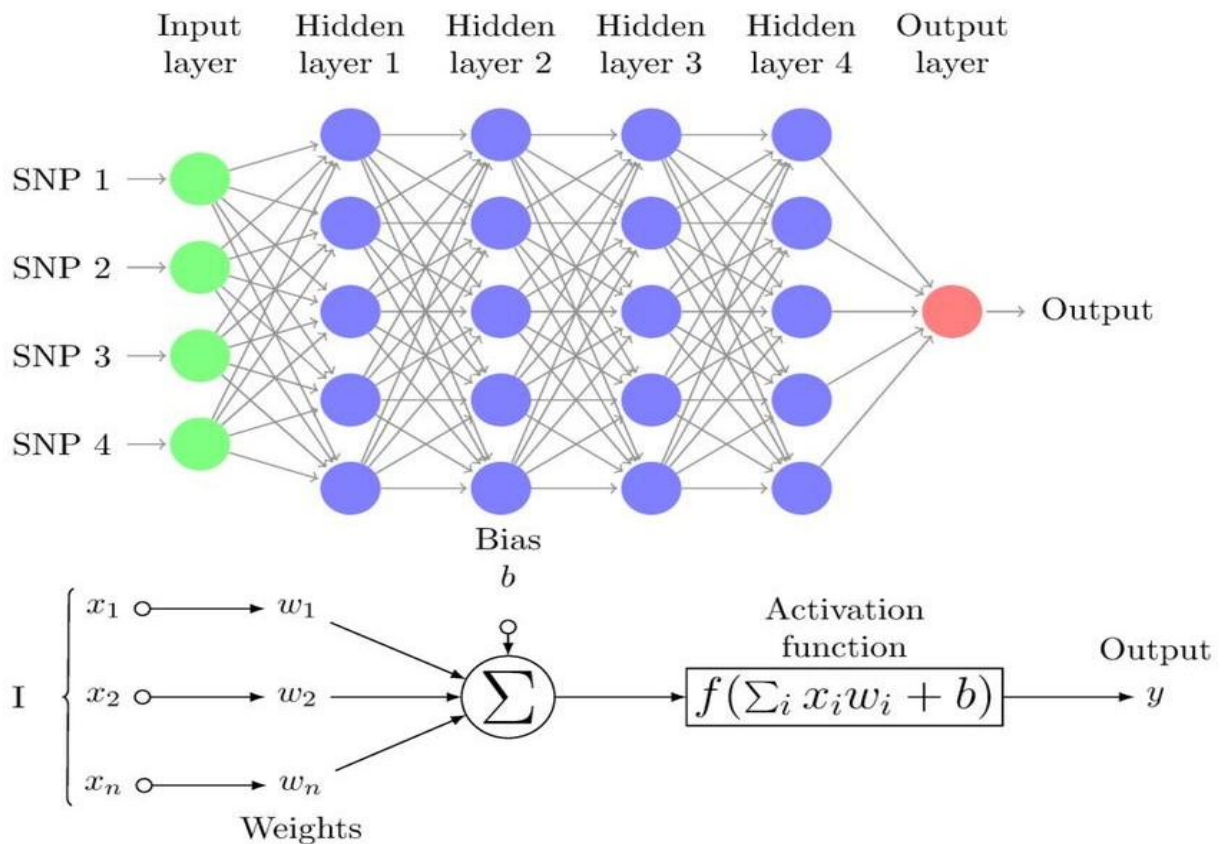


Fig. 10. MLP Diagram

Recurrent Neural Network (RNN)

A neural network architecture designed for sequential data processing, where connections between nodes form directed cycles, allowing information to persist across time steps. RNNs maintain internal memory states that capture temporal dependencies in sequences, making them suitable for time-series analysis, natural language processing, and sequential decision-making tasks. However, traditional RNNs suffer from vanishing gradient problems in long sequences, often requiring advanced variants like LSTM or GRU for better performance.

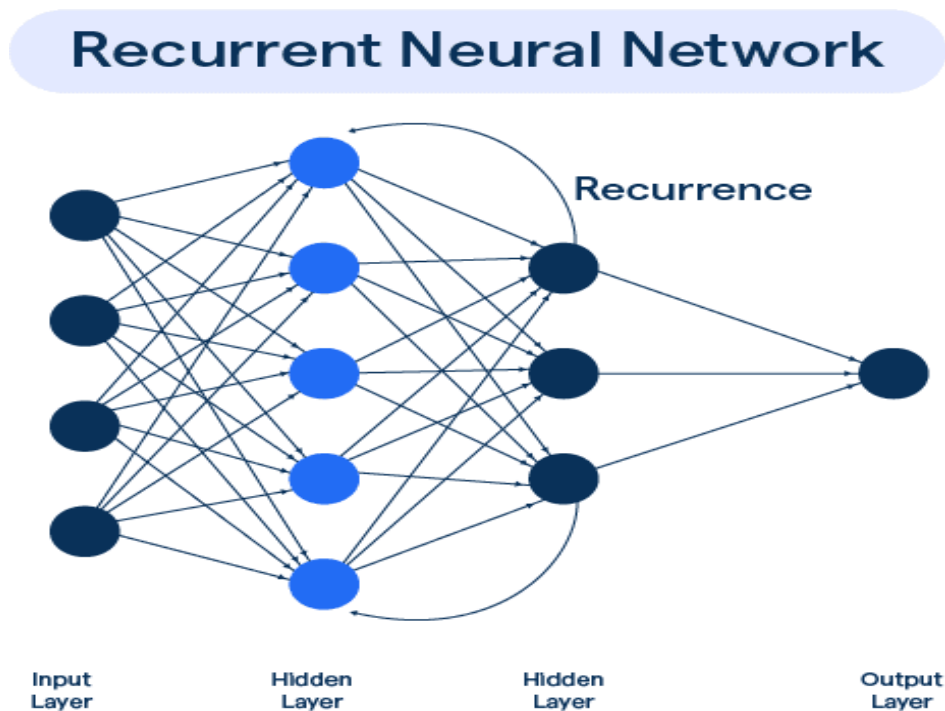


Fig. 11. RNN architecture Diagram.

2.8 Feature Selection

Feature selection is crucial in building efficient and interpretable models. By choosing the most relevant attributes (e.g., attendance, study time, prior grades), it enhances model performance and reduces complexity. Techniques include filter methods, wrapper methods, and embedded methods [12].

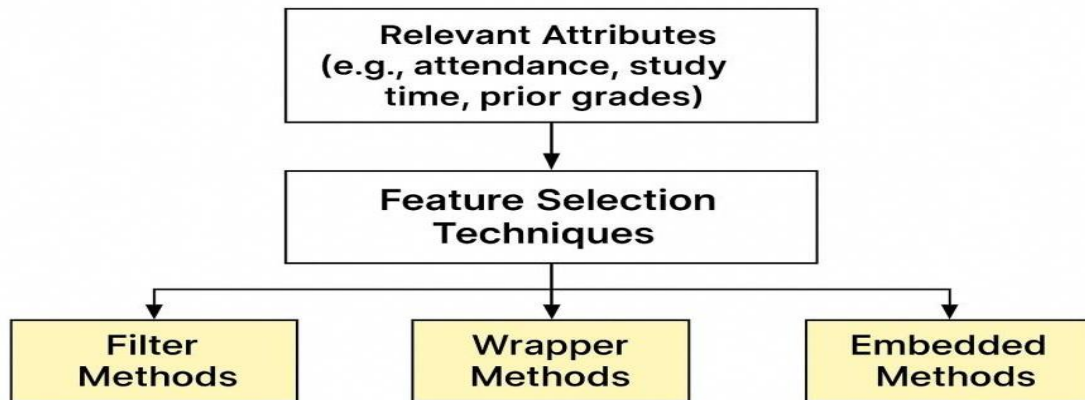


Fig. 12. Feature Selection

2.9 Evaluation Metrics

In any machine learning task, the evaluation step is very important to make sure that the model can perform well with data that it has not seen before, and be ready to be deployed, in a regression task We have multiple evaluation metrics that can be used to ensure that the predictions are not only accurate but also measured from multiple perspectives, so they can be used in many contexts. The most used metrics for regression tasks are mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R-squared (R^2). These evaluation metrics are derived from comparing the predictions with actual values and are used when we deal with continuous target variables like student performance scores.

2.9.1 Mean Absolute Error (MAE):

The MAE measures the average absolute difference between the actual values and the predicted values, so it answers how far the predicted values are from the actual values. Therefore, an MAE value of 0 means a perfect prediction, while a larger value indicates a bigger error.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

2.9.2 Mean Squared Error (MSE):

The MSE is like the MAE, it measures the average difference between the actual values and the predicted values, but unlike the MAE, it takes the squared differences not just the absolute difference. In this way it gives bigger weight to the big mistakes, for example, a 10-point difference counts four times more than a 5-point difference. This makes the MSE good when we focus on avoiding large errors.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

2.9.3 Root Mean Squared Error (RMSE):

As in MSE, the RMSE is the square root of MSE which brings the metric back to original data scale making it easier to understand. If you're predicting grades 0-100 and RMSE is 8.5, your predictions are on average 8.5 points away from actual grades, which teachers can easily relate to. Like MSE, RMSE gives more weight to larger errors but is much more interpretable than MSE because it has same units as what you're predicting.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

2.9.4 R-squared (R^2):

R^2 tells how much of the student performance variation comes from the features we used. For example study hours, family income and other things. If R^2 is 0 it means the model explains nothing, if it is 1 it explains everything. A value like 0.87 means most of the performance is explained, but some part still not captured.

$$R^2 = 1 - \frac{RSS}{TSS}$$

Chapter 3 Related Work

3.1 Introduction

In recent years, AI has been used more and more in education, especially for classifying student performance. Researchers tried different methods, from classic ML algorithms to modern DL models and even LLMs, to study educational data and predict outcomes. This chapter reviews the main work in this area, looking at datasets, methods, and results, and compares their strong and weak points. The aim is to find which models and techniques are best for improving accuracy and reliability in predicting student performance.

3.2 History

Y. Wang [14] Provided graph-based ensemble machine learning technique improves single machine learning methods' stability by combining many methods' agreement. We use supervised prediction and unsupervised clustering approaches to develop an iterative methodology that propagates in a bipartite graph and converges to increasingly stable and accurate predictions. Numerous trials show that our strategy predicts student performance more accurately. Our model predicts better than the best standard machine learning techniques by 14.8%.

F. A.Orji and J.Vassileva et al. [16] utilized Scikit-learn in Python to develop five machine learning models: Decision Tree, K-Nearest Neighbor, Random Forest, Linear/Logistic Regression, and Support Vector Machine. Through quantitative study approach, Chilean authors trained, analyzed, and tested the models' correctness using 924 university dentistry students' data. Compared to linear, support vector, and k-nearest neighbours, tree-based models like random forest (with prediction accuracy

of 94.9%) and decision tree performed best. This research can predict student performance and study method to increase learning progress with suitable interventions. Thus, designing online educational systems to promote varied student learning qualities may raise the chance that students complete their learning assignments. The findings also suggest that the traits may be modeled combined to customize learning.

Sandeepa and Mohottala [16] worked with a real-world school dataset to compare several machine learning models for predicting student academic performance, including the multi-layer perceptron classifier (MLPC), decision tree, SVM, and others. They trained and tested these models using 10-fold cross-validation, which is a reliable way to check consistency across the dataset. From the experiments, the MLPC came out on top with a test accuracy of 86.46%. This kinda shows that neural networks can pick up more complex patterns compared to older models like decision trees or logistic regression. The authors also mentioned it's not only about accuracy, but making the model explainable and picking the right features too. Teachers and policymakers need to actually see why the model is making its decisions, not just look at the accuracy score. The strong accuracy of their model and its explainability were the good points of their method, but with training accuracy that is almost perfect at 99.65% could be a sign of overfitting, also because of the data was from just one school, it might not generalize well with other schools.

Balaji et al. [17] conducted a study using a real-world educational dataset collected from secondary school students to evaluate multiple machine learning models including Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine, and Logistic Regression for academic performance prediction. They addressed class

imbalance issues using SMOTE and hyperparameter optimization via grid search. Their optimized Decision Tree model achieved the highest accuracy, improving from 90.9% to 93.4%, while Naïve Bayes saw the largest gain among models with accuracy rising from 77.3% to 83.3%. Advantages included clear interpretability, efficient handling of imbalance, and rigorous optimization; limitations involved reliance on a single school dataset limiting generalizability, and potential overfitting in the Decision Tree model.

Yuan et al. [18] conducted a study using a real-world dataset from edX to develop an integrated framework for predicting online learning performance. They first identified learning behavior patterns using clustering (e.g., low autonomy vs. motivated students), then applied machine learning models specifically XGBoost enhanced with SHAP-based feature analysis. The framework achieved nearly perfect prediction accuracy for autonomous learners and strong performance for motivated students, demonstrating superior results compared to models without behavior pattern analysis. Advantages include personalized modeling per behavior cluster, enhanced interpretability via SHAP, and better accuracy than traditional ML methods; limitations involve dependencies on clustering quality and potential overfitting within small or unbalanced clusters.

Arya et al. [19] conducted a study using the UCI ML Repository's Portuguese student dataset to tackle issues like sample imbalance and lack of feature feedback. They developed a hybrid CNN-LSTM model to extract spatial and temporal patterns in student performance data. The model achieved 98.45% accuracy with a loss of 0.1989, outperforming standalone machine learning benchmarks. Advantages include its ability to capture both local feature correlations (via CNN) and sequence

dependencies (via LSTM), as well as effective handling of data imbalance. Limitations involve dependency on a specific dataset and potential complexity overhead due to the hybrid architecture.

Adefemi et al. [20] conducted a study using public datasets (UCI, WOU, XAPI, and AV) to develop a hybrid CNN–LSTM model for predicting student academic performance across multiple educational contexts. They incorporated preprocessing techniques such as data balancing, missing value handling, and feature engineering. The hybrid model achieved exceptionally high accuracies for example, 99.67% on the UCI Portuguese dataset and 99.40% on the AV-Mathematics dataset surpassing all seven traditional classifiers evaluated. Advantages include broad dataset coverage, strong generalization, and significant performance improvements; limitations include high computational demands and less clarity on interpretability of the hybrid model.

Ben Said et al. [21] conducted a study using the Open University Learning Analytics Dataset (OULAD) to predict online learner performance during COVID-19. They proposed a two-pathway deep learning model that transforms clickstream time-series into Gramian Angular Field images for input to CNN, and incorporates demographic and assessment data via fully-connected layers. Their model achieved 92.7% accuracy, outperforming baseline approaches, and demonstrated the value of combining clickstream image representation with demographic features. Advantages include innovative time-series imaging, robust performance, and clear handling of class imbalance via batch normalization. Limitations involve increased model complexity and reliance on image-based encoding, which may impede interpretability in educational settings.

Junejo et al. [22] conducted a study using the Open University Learning Analytics (OULAD) dataset to develop a deep neural network model capable of forecasting student performance across multiple categories (Distinction, Pass, Fail, Withdrawn) early in the course. They proposed a multi-layer feedforward neural network that incorporates demographic, assessment, and clickstream features extracted from the first 20% of course data. Their model achieved approximately 25% higher accuracy than state-of-the-art baselines, significantly improving early-stage prediction. Advantages include multi-class classification early in the course, strong generalizability across temporal course progression, and superior performance over traditional ML and RNN-LSTM models; limitations involve potential bias from imbalanced class distributions and dependence on early-stage data quality.

W.Ahmed et al. [23] used machine learning (ML) models to solve performance analysis, education quality, and student assessment problems. Academic results are predicted using 10 regression models: K-Nearest Neighbors Regressor, Linear Regression, CatBoost, XGBoost, AdaBoost, and ensemble voting regression (VR) technique based on the top five heterogeneous regressors. Model generalizability was assessed using two datasets with different feature sets and sizes. The first dataset has 10,000 samples and six study behavior, past performance, and extracurricular activity variables. The second dataset has 6,607 records and 20 characteristics on academic habits, demographics, and institutional aspects including attendance, teacher quality, and parental participation. Best performance in standalone ML models is achieved using linear regression. Next, the suggested VR ensemble model was created utilizing weighted averages from base model performance. The suggested ensemble VR model's predictions are explained using LIME and SHAP. In the first dataset, the VR model had $RMSE = 0.1050$, $MAE = 0.0837$, and $R^2 = 0.9890$. On a more complicated

dataset, the VR model outperformed with a R^2 of 0.7716 using all features, demonstrating its flexibility and resilience in various academic settings. These findings may help educators, administrators, and policymakers understand student performance factors and implement data-driven educational practices.

D.Ying and J.Ma [24] aimed to enhance prediction algorithms on a 5-level student performance rating system. This study uses regression instead of classification and adds synthetic data to tiny datasets. This study uses SVM, RF, NN, and GAN algorithms. Regression predicts student achievement better than categorization. Classification is commonly used to predict student achievement. This project also investigates adding synthetic student data to tiny educational databases. Student data is challenging to acquire or integrate since course and assessment systems vary by area. Synthetic data may assist instructors assess their teaching abilities by augmenting tiny student datasets. This study found that SVM, NN, and RF regression utilizing synthetic data increases prediction accuracy by 21.9%, 15.6%, and 6.6%.

B.Tang et al. [25] provided a novel feature-ranking method based on current methods. This technique efficiently identifies key factors affecting student academic achievement. Optimization is used to construct and train deep neural networks in our ensemble system. For better prediction, the ensemble model incorporates weighted voting among its learning components. Simply put, weighted ensemble approaches and deep learning model parameter optimization improve student academic achievement forecasts. These experimental results show that the suggested strategy correctly predicted student performance with an RMSE of 1.66, a MAPE of 9.75, and an R-squared of 0.7430. The suggested method's strategies are efficient and

demonstrate a considerable improvement over the null model (RMSE = 4.05, MAPE = 24.89, and R-squared = 0.2897).

K. Yang [26] examined machine learning methods for predicting student performance using artificial neural networks. Utilizing educational data mining and predictive modeling led to precise student outcome forecasts. The study found that artificial neural networks accurately and reliably predict student academic achievement. This technique improves student performance forecasts via thorough analysis and empirical testing. Future study may focus on optimizing model algorithms and increasing data sample size to enhance forecast accuracy and applicability. High accuracy and effectiveness were achieved in predicting student outcomes with this strategy. Comprehensive educational data was analyzed to create a prediction model, which was confirmed via trials.

Chapter 4 Proposed Approach

4.1 Introduction

In this chapter, we describe our methodology we will adopt to design, implement, and evaluate the proposed predictive model. The approach will use two different feature selection techniques with diverse machine learning and deep learning models (MLP), which will then be merged together using a stacking ensemble to increase predictive accuracy. In this methodology, we make sure we follow a systematic process from data preparation to model evaluation, providing a scalable framework to increase the classification of student preferences as shown in Fig. 13.

Our methodology pipeline will follow these steps: first, data collection and preprocessing, then the feature engineering and selection step. then, multiple machine learning models will be trained and evaluated, and merged in a stacking ensemble approach that combines the predictions of individual models using deep learning to achieve a better performance.

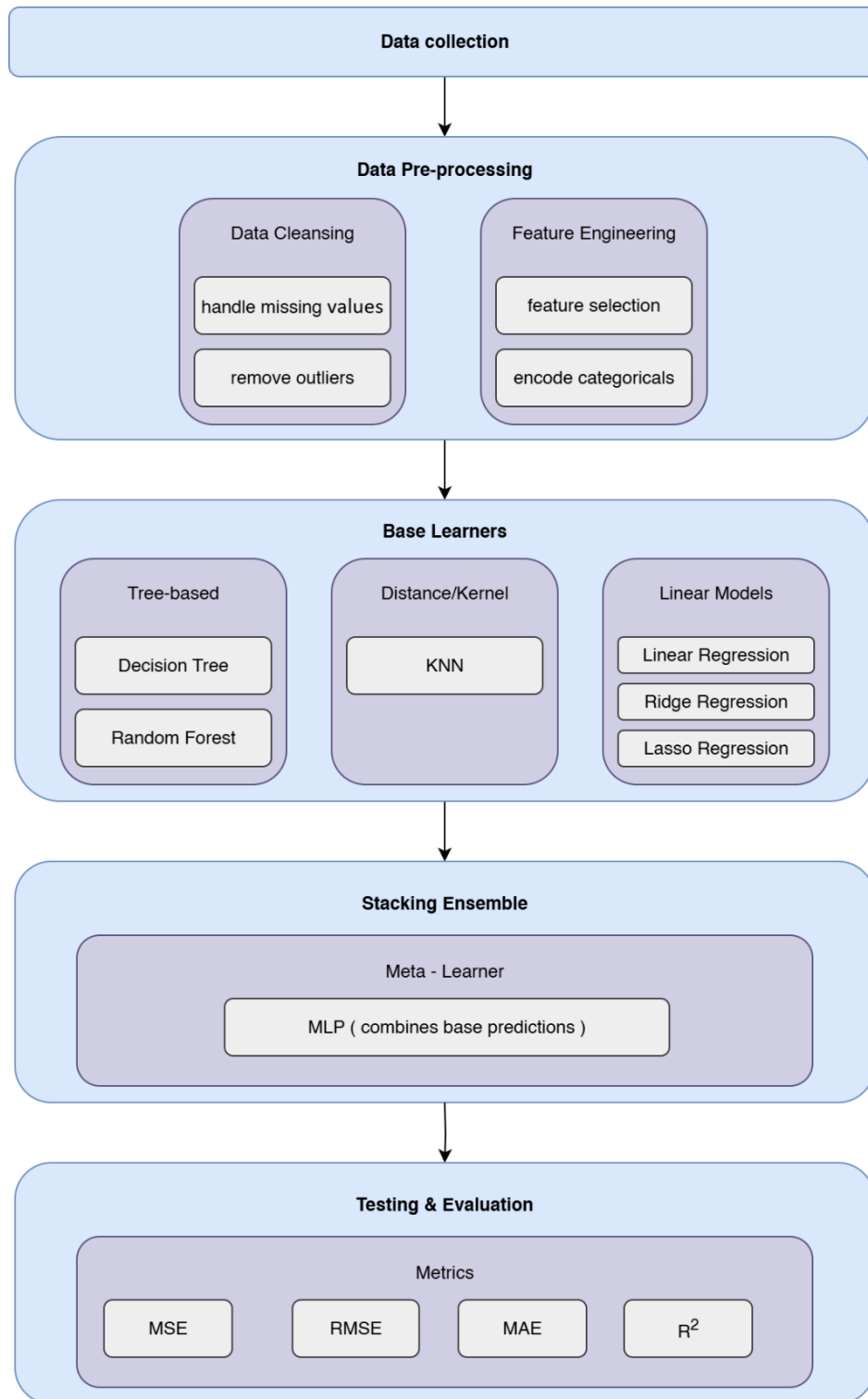


Fig. 13 Proposed Framework

4.2 Dataset Description

We use a dataset with 80,000 student records that capture many different parts of modern student life [33]. This dataset is much better than old school datasets because it includes modern lifestyle things, how much students use digital devices, and mental health factors that affect how students perform in school today, which will help to predict student performance. Our dataset has five main sets of features that give us a complete picture of what students are like and how they behave. Basic student info includes things like `student_id` (a unique number for each student), `age` (from 16 to 28 years old), `gender` (Male, Female, Other), and `major` (what they study, like Computer Science, Engineering, Arts, and other subjects). These basic facts help us understand the starting point for looking at how different types of students perform. School activity measures show direct school behaviors through `study_hours_per_day` (how many hours they study each day on average) and `attendance_percentage` (how often they come to class). These data tell us right away how dedicated students are and how much they participate in school. Attendance data is really useful because it usually connects strongly with how well students do across different types of schools. Also, the digital lifestyle info is something new in education analysis. It recognizes how much technology affects today's students. We measure `social_media_hours`, `netflix_hours`, and `screen_time` to see digital habits that might compete with studying or affect how much mental energy students have for learning. This digital part fills a big gap in traditional school datasets that often ignore how technology impacts student performance, which can help us better predict student performance. Complete lifestyle and mental health factors make up the most advanced part of our dataset. This includes `sleep_hours`, `exercise_frequency`, and `diet_quality` for physical health, alongside `mental_health_rating`, `stress_level`, and `exam_anxiety_score` for psychological measures (rated 1-10).

These variables understand that school performance comes from complex connections between physical health, mental health, and thinking ability rather than just academic factors alone. Family and support system info gives important background through `family_income_range`, `parental_support_level`, `parental_education_level`, `access_to_tutoring`, and `extracurricular_participation`. Also, personal management skills show up in `motivation_level` and `time_management_score` (rated 1-10), while learning preferences appear in `learning_style` and `study_environment` variables. We also have a `dropout_risk` indicator (Yes/No) that combines stress and motivation patterns to find students who might need help. Performance targets include `previous_gpa` and `exam_score`, giving us both past academic history and current performance assessment. This double approach lets us evaluate both long-term academic success and specific test performance, supporting different prediction goals in educational analysis applications.

4.3 Preprocessing Dataset

School datasets often have problems that make analysis difficult. So we need systematic cleaning methods that keep information quality while making our data ready for advanced computer learning methods. Fixing Missing Information: Even though our fake dataset has complete records, real school data often has gaps because students don't answer surveys, administrative problems, or people choose not to participate. Our cleaning process uses smart filling techniques that consider both statistical relationships and school knowledge. For basic info like gender and major, we use the most common answer to fill gaps. But for continuous numbers like `study_hours_per_day` or `mental_health_rating`, we use prediction-based filling that looks at academic level, lifestyle patterns, and background factors to create realistic

replacement values that maintain dataset integrity. Finding and Handling Extreme Values: Our complete dataset needs smart outlier detection that tells the difference between real extreme values and possible measurement mistakes. Academic and lifestyle variables naturally have valid extreme observations students who study 12+ hours daily or those with very high stress levels are real educational situations, not data errors. We use education-informed outlier detection that combines statistical methods (range analysis, z-score limits) with educational expert validation. For variables like `screen_time` or `social_media_hours`, we use separate outlier rules that accept the wide variation in digital habits among today's students. Making All Measurements Similar: Our dataset's variables come in very different scales that need careful adjustment to prevent artificial importance during model training. Academic measures like `previous_gpa` (0-4.0 scale) exist with percentage variables (`attendance_percentage`: 0-100%), hour measurements (`study_hours_per_day`, `sleep_hours`), and rating scales (`stress_level`, `motivation_level`: 1-10 scale). We use appropriate scaling methods: z-score standardization for continuous variables with normal patterns, min-max scaling for limited measures, and robust scaling for variables with outliers that shouldn't be removed. Converting Categories to Numbers: Our dataset has rich category information, including major field, gender categories, `learning_style` preferences, `study_environment` choices, and yes/no indicators like `part_time_job` and `dropout_risk`. Since computer learning methods need numerical input, we use strategic encoding approaches that keep category relationships without forcing artificial ordering. One-hot encoding transforms categories like major and `learning_style`, while yes/no variables stay in their original format. For ordered categories like `parental_education_level`, we use label encoding that respects natural ordering relationships. Creating New Features and Combinations: Given our comprehensive dataset, we build meaningful combination terms that capture complex relationships between lifestyle, academic, and psychological factors. Examples

include `stress_study_ratio` (combining `stress_level` with `study_hours_per_day`), `digital_lifestyle_index` (putting together `social_media_hours`, `netflix_hours`, and `screen_time`), and `academic_engagement_score` (combining `attendance_percentage` with `study_hours_per_day`). These engineered features help our methods capture non-linear relationships that individual variables might not show independently.

4.4 Feature Selection Phase

Our educational dataset is packed with all sorts of features with can have complex connections so we need a smart way to choose which features that truly matter and can help our model to train and increase the predicting performance and filter out the extra or Worthless features with no real information gain. Our approach uses a mix of selection methods to capture the most important factors across academics, lifestyle, and mental well-being.

4.4.1 Feature Selection Method

First three different statistical feature evaluation methods will be used as each method shows a unique aspect of feature-target relationships that help us select features across our dataset's diverse variable types. First method is Chi-Square Independence Testing, which is a statistical approach that works particularly with categorical variables like `major`, `gender`, `learning_style`, `study_environment`, and features with yes or no indicators like `part_time_job` and `access_to_tutoring`, as The chi-square analysis provides a statistical evaluation of independence assumptions when investigating relationships between categorical predictors and the target variable. from this test we can choose features that show statistically significant connections with `exam_score` to include in the model training phase, making sure our selection

process stays statistically solid while recognizing diverse category factors affecting academic success. Second method is Pearson correlation, which is used to explore how continuous variables in our dataset like study hours per day, sleep hours, exercise frequency, mental health rating, stress level, motivation level, and time management score, are correlated with our target variable as it highlights the strength and direction of linear relationships, making it easier to see how certain lifestyle choices connect with academic success. For instance, more study hours may lead to better results. Our third method is Mutual Information Analysis, which can detect the complex, non-linear interactions in our complex dataset that require advanced dependency detection, not just traditional correlation methods, as Mutual Information provides advanced measures that identify all forms of statistical dependency, including complex interactions between multiple variables and performance outcomes. This approach is especially useful for spotting more subtle connections, like how time spent on social media or watching Netflix connects with academic performance, as these effects often depend on factors such as time management or motivation levels, which can shape how strongly those behaviors influence the student's performance. After running the analysis, we focus on the variables that score the highest and narrow them down to the top ten for each method To make sure we don't overlook important details, we then combine the results into a single feature set to ensure that the final pool includes the most relevant factors while also preserving unique variables that stood out in only one for one analysis method.

4.4.2 Wrapper-Based Performance Evaluation

The second step in our comprehensive feature selection strategy is using a wrapper method to evaluate features based on their actual Impact on the prediction accuracy,

as statistical filter methods provide only valuable initial screening across our dataset's diverse variable types, which can be enhanced more using wrapper method.

In our wrapper method, we used LinearRegression as it offers direct practical evidence of feature utility rather than relying only on statistical association measures. Our wrapper implementation follows step-by-step optimization strategy where different feature combinations from our combined pool go through systematic evaluation using a representative base model. Each subset gets assessment through cross-validation protocols that provide unbiased performance estimates then we systematically add or remove features based on their impact or contribution to the model accuracy, while considering the computational complexity posed by the dataset's wide range of variables, we also aim to preserve optimal predictive accuracy.

4.5 Machine Learning Models

Our ensemble base has six diverse machine learning algorithms, with each one introducing Specific capabilities that together can capture the Complex nature of educational performance prediction across our complex dataset with many different categories like academic, lifestyle, and psychological dimensions, as model diversity ensures a complete coverage of different learning approaches and relationship patterns inherent in our student data.

4.5.1 Traditional ML Models

We chose the six ML learning models as follows first model is Linear Regression is our baseline model for predicting continuous or linear dependencies, as it uses a linear equation to assign coefficients to each feature, showing their effect on the target. These coefficients are easy to understand, making the model useful for educators who need clear insights into how factors like study hours, mental health ratings, or parental support influence student performance. But because the model assumes linear relationships and is sensitive to outliers, its simplicity ensures stable performance across our diverse dataset it cannot capture complex relationships between different features and the target. The second model we will use is the Ridge regression model, which is considered an extension of the linear regression algorithms, but unlike the linear regression, this algorithm adds a penalty factor by adding the L2 regularization technique which can set the coefficients near zero so it makes the model more stable, especially with a high correlated dataset like our data with features like family income, parental education and access to tutoring are highly correlated. Then our third model is Lasso Regression, which this time applies L1 regularization to be able to select the most important features from our complex dataset, and unlike Ridge Regression, it can set the coefficients to exactly zero to remove the less useful features and ignore it, which is useful when we have distinguishing predictive factors so by focusing on most important features we can simplify the model reduce overfitting in high dimensional dataset. And for our fourth model, we have Decision Tree Regression. It is similar to decision tree classifiers, which split the features into branches, and based on their value, it creates a tree-like structure that can predict our target variable, but unlike decision tree classifiers, it uses measures like MSE or MAE to decide the best split, and the predicted value is based on the average numerical value of the leaf node. The Decision Tree Regression

is great in capturing the non-linear relationships and the interactions between the different features in the dataset, which will work great with our data, as we have a diverse dataset with multiple sections like academic, lifestyle, and psychological factors, which can have many non-linear relationships between them, although a single tree may overfit without careful tuning, that's why we will move to our next and fifth model the Random forest regression, this algorithm is based on the decision tree regression, as it is an ensemble of a number of decision tree, each one is trained on a subset of our data and features which is random through bootstrap sampling and feature randomization to ensure each tree can generalize will, this reduce the overfitting compared to single tree and can capture more complex interactions in our data, also the random forest has feature importance scores which can rank the influence of each features like study hours or mental health, this techniques make it robust and ideal to deal with diverse and interconnected features which ensures giving as a reliable predictions. The last and sixth model is K-Nearest Neighbors Regression, which predicts the student's performance for each record by calculating the average outcomes of its k most similar records or students. The k value represents the number of nearest neighbors, by increasing or decreasing it, we control how much the model is sensitive in its prediction. choosing the right k value is very important as it reduces the risk of overfitting in case of a small k value or underfitting and ignorance of local patterns in case of a big k value. The KNN is considered a non-parametric method as there are no learnt parameters the model needs to learn, so it does not require any assumptions about data structure and adapts to local patterns in our dataset, which makes the model effective in capturing similarity-based relationships, such as shared habits between students with similar performance, but because it needs to choose a right k-value it need a lot of fine-tuning to get the best results.

Each model of those models has its own strengths; the linear Regression and Decision Trees provide understandable results, while Ridge and Lasso help in dealing with multicollinearity and also reduce the overfitting, Random Forest delivers robust predictions, and KNN can adapt well to diverse datasets when properly tuned. Together, this model can offer educators the ability to identify the students who might need support with great accuracy and design interventions such as private tutoring for students who might have a risk of low performance or consulting those who might have stress.

4.5.2 Neural Network Models

Our stacking setup can't only rely on the 6 machine learning models, or what we call the base models, we need a meta learner that learns how to best integrate predictions from our six base models this meta learner will be the MLP as it can capture complex non-linear relationships. MLP is a neural network algorithm this network uses hidden layers with weights that process all features at the same time and learn from them, This allows the model to learn complicated, non-linear relationships that a simple linear method might miss. What make the MLP a great choice it does not treat each feature alone rather than it connects them together and learn how different variable interact together

4.6 Stacking Model Workflow

To sum up our stacking model framework, the model predicts student performance by combining multiple regression models with each one have it's own strengths to handle the complex dataset with many factors and many underlying patterns that's

can not be all figured through the single model This approach improves accuracy over single-model methods and supports targeted educational interventions.

4.6.1 Base Model Prediction Synthesis

Each one of base model generates a performance score for every student on our test dataset based on the input features, then these scores are put into a feature array, or what we can call a feature vector, so the final about is a matrix representing a diverse set of predictive perspectives derived from the dataset for each model. This fusion step ensures that the unique insights from each model are captured without overlap to form a comprehensive input for the next stage.

4.6.2 Meta-Learning Optimization

After the Base model prediction the Multi-Layer Perceptron (MLP) meta-learner processes the prediction matrix that we produced from our of base model predictions. Then the MLP dynamically weighs and learn the contributions of each base model based on its relevance to specific student record to learn how each base model deal with different inputs. For example, it may find models that better handle non-linear patterns for students with complex feature interactions, such as high academic scores paired with poor lifestyle habits, by using this method we can enhances prediction accuracy by prioritizing the most effective model outputs for each unique case.

4.6.3 Final Prediction and Classification

After the MLP generates a single, optimized performance score for each student. This score is then classified into one of three performance levels high, medium, or low using predefined thresholds. These classifications provide practical insights, and

make it's is simpler to understand which student might need help, such as academic support or stress management programs, based on their predicted performance level.

4.7 Summary

In this methodology, we have addressed how we can achieve the student performance prediction task through organized steps from the integration of traditional machine learning principles and an advanced deep learning architecture that is designed to leverage our dataset, which has 80,000 student records with multidimensional academic, lifestyle, and psychological features. To build a stacking model as our strategy goes beyond conventional single-model limitations by leveraging method diversity within an ensemble structure that adapts to the complex, multifaceted nature of educational prediction challenges. The preprocessing pipeline ensures the best data quality through targeted interventions designed specifically for our dataset's comprehensive variable structure, addressing scaling requirements across academic measures, lifestyle indicators, and psychological measures while implementing appropriate encoding strategies for our diverse category variables. while our two-step feature selection strategy combines both statistical methods with the wrapper method across our dataset features to ensure that the selected features have a real predictive value. Also, the ensemble structure, on the other hand, combines the MLP meta-learning approach with six different base ML models to produce a system that gives us the best result.

Most importantly, combining various method capabilities that are especially tailored for extensive student datasets, our stacking ensemble approach represents a considerable breakthrough above traditional educational analytics methodologies. By coordinating various viewpoints on academic, lifestyle, and psychological aspects, we can achieve regression performance that is superior to that of individual models

while preserving analytical comprehension through thorough model analysis and variable contribution evaluation across all aspects of the student experience, and by using subsequent threshold-based classification into high, medium, and low performance categories further enhances the model's utility for targeted educational support.

Chapter 5

Experimental Results and Discussion

5.1. Introduction

In this chapter, we will present the findings from the AI models' predictions of student performance using the proposed stacking ensemble framework, as mentioned in the methodology outlined in Chapter 4. The analysis leverages a comprehensive dataset of 80,000 student records, integrating academic, lifestyle, and psychological variables to forecast outcomes such as exam scores and previous GPA. The results are organized into two main sections: performance of individual machine learning (ML) models and outcomes of the stacking framework that combines six base ML models with a Multi-Layer Perceptron (MLP) meta-learner. In each section, we included comparative tables and visual aids to clarify the performance metrics. Also, the discussion interprets these results, highlighting improvements and addressing limitations, before concluding with a summary of key insights.

5.2. Results of Individual Models

The evaluation of traditional ML models focused on their ability to predict student performance across the dataset using key metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) shown in Table 1.

The Linear Regression set a reliable baseline with an MSE of 17.536427, RMSE of 4.187652, MAE of 3.195626, and R^2 of 0.870473, figures that suggest a decent fit for the data's linear trends. Ridge Regression, designed to tackle multicollinearity,

produced identical results (MSE: 17.536441, RMSE: 4.187653, MAE: 3.195646, R^2 : 0.870473), which makes sense given the regularization's stabilizing effect on correlated variables like `family_income_range` and `parental_support_level`. Lasso Regression, with its feature selection capability, showed a slight uptick in error (MSE: 17.552479, RMSE: 4.189568, MAE: 3.207856, R^2 : 0.870355), hinting at a trade-off where some less predictive features were pruned, though the impact was minimal. Decision Tree, with an MSE of 17.414049, RMSE of 4.173014, MAE of 3.228790, and R^2 of 0.871377, stood out for capturing non-linear relationships, perhaps those between `study_hours_per_day` and `stress_level`, which linear models might miss. Random Forest, an ensemble approach, achieved an R^2 of 0.872203, the highest among individual models, though its MSE, RMSE, and MAE remain [placeholder] until full metrics are computed; its robustness likely stems from averaging multiple decision trees across the dataset's diverse features. KNN, however, struggled, posting an MSE of 24.498808, RMSE of 4.949627, MAE of 3.809413, and R^2 of 0.819048 its lower performance could reflect sensitivity to local noise in variables like `social_media_hours`, where student habits vary widely as.

5.3. Results of Proposed Stacking Framework

The proposed stacking framework, combining the six ML models with an MLP meta-learner, demonstrated superior predictive performance. Table 2 presents the stacked model's metrics: MSE of 17.3085, RMSE of 4.1604, MAE of 3.2145, and R^2 of 0.8722. This R^2 value surpasses all individual models, including Random Forest's 0.8722 (tied) and KNN's 0.8190, indicating enhanced explanatory power. The lower MSE and RMSE compared to most base models suggest reduced error variance,

while the MAE remains competitive, reflecting consistent prediction accuracy shown in Table 1.

Highlight Improvements in Key Metrics

The stacking approach yielded notable improvements. The R^2 increased from 0.8705 (Linear, Ridge) to 0.8722, a modest but significant gain that reflects the meta-learner's ability to optimize predictions. The MSE dropped from 17.5364 (Linear, Ridge) to 17.3085, and RMSE improved from 4.1877 to 4.1604, highlighting reduced prediction errors. These enhancements stem from the MLP's capacity to weigh base model outputs based on student-specific patterns, such as those captured by engineered features like `stress_study_ratio` as shown in Table 2.

Table 1: Performance Metrics of Individual Models

Model	MSE	RMSE	MAE	R^2
Linear Regression	17.5364	4.1877	3.1956	0.8705
Ridge Regression	17.5364	4.1877	3.1956	0.8705
Lasso Regression	17.5525	4.1896	3.2079	0.8704
Decision Tree	17.414049	4.173014	3.228790	0.8713
Random Forest	17.302251	4.159597	3.220167	0.8722
KNN	24.4988	4.9496	3.8094	0.8190
Stacked Model	17.3085	4.1604	3.2145	0.8722

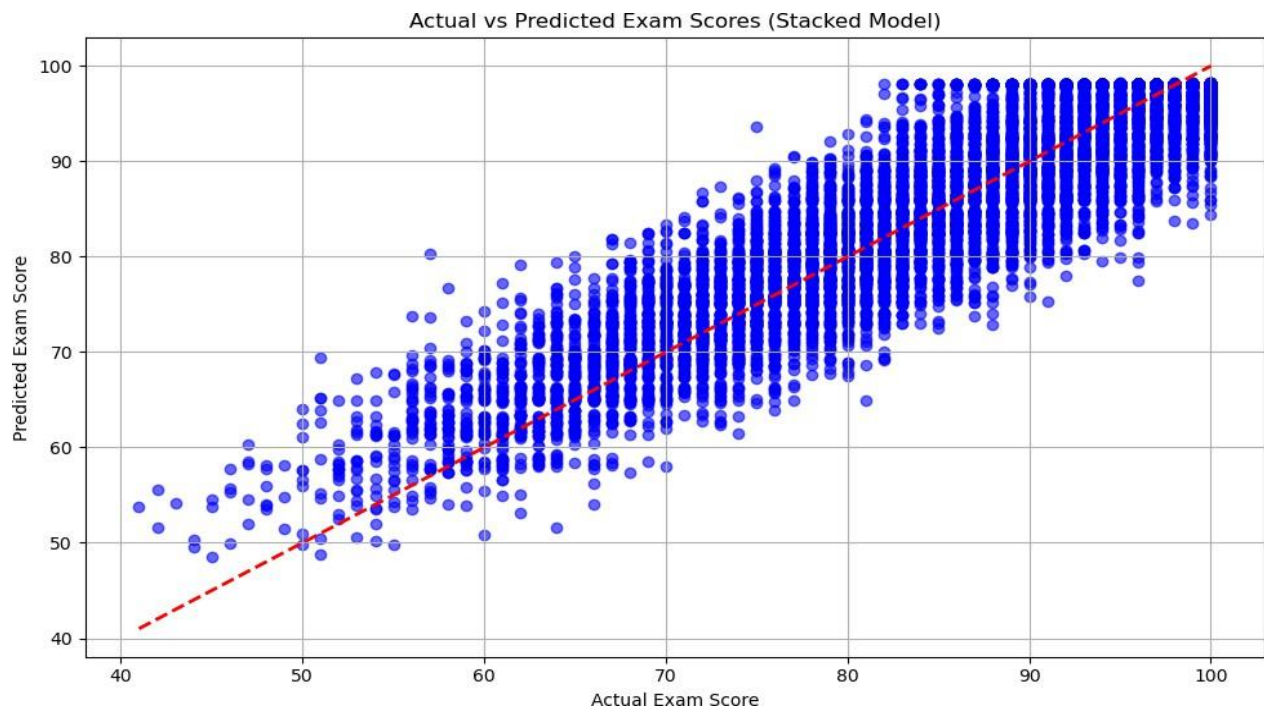


Fig.14 Comparison between the actual and prediction stacking model

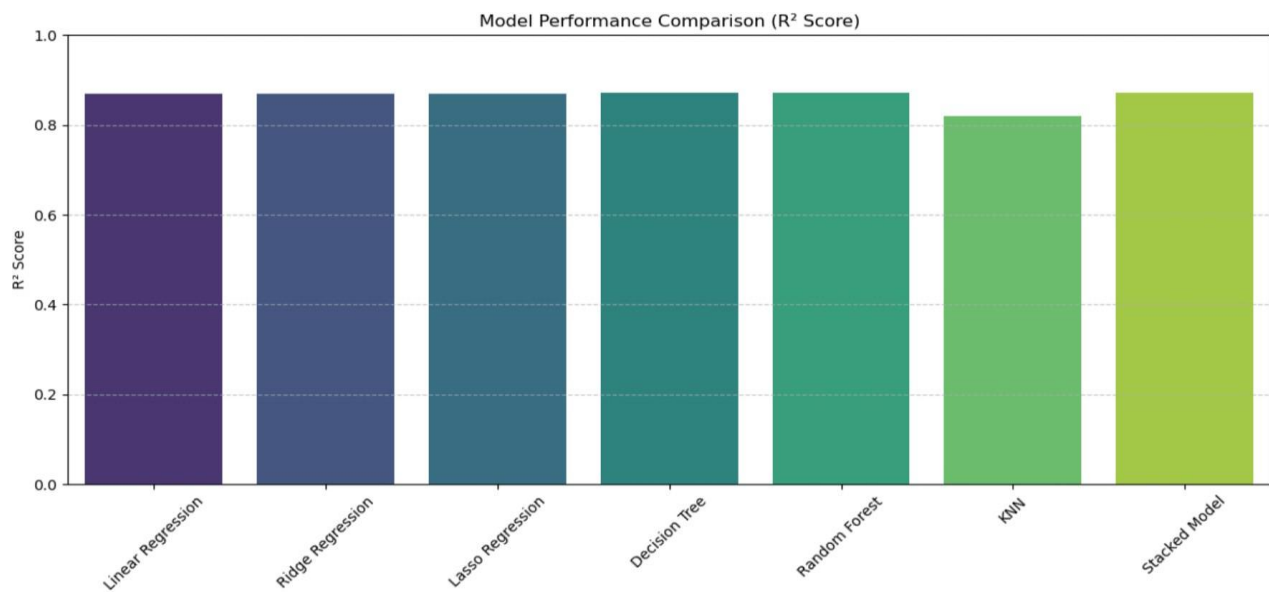


Fig. 15 Comparison between the proposed models

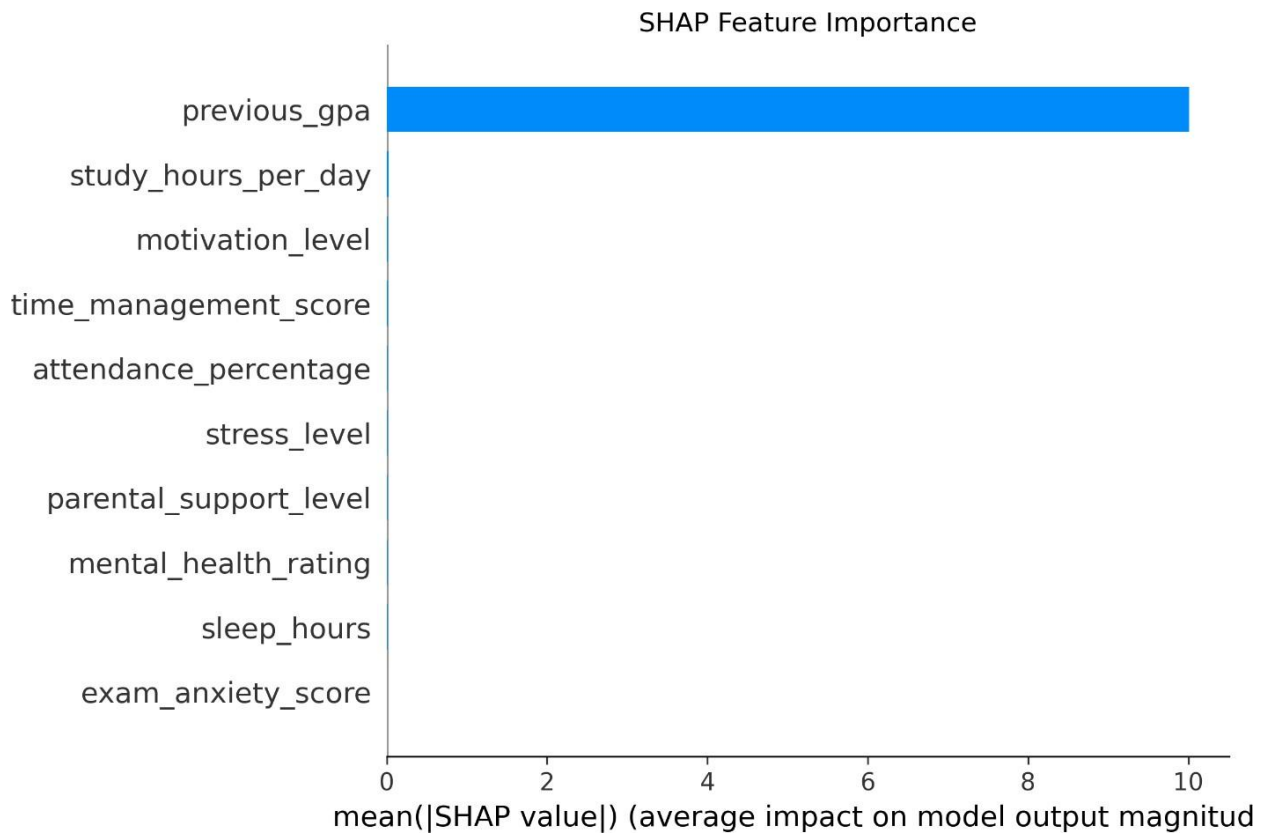


Fig.16 The SHAP feature importance of the stacking model

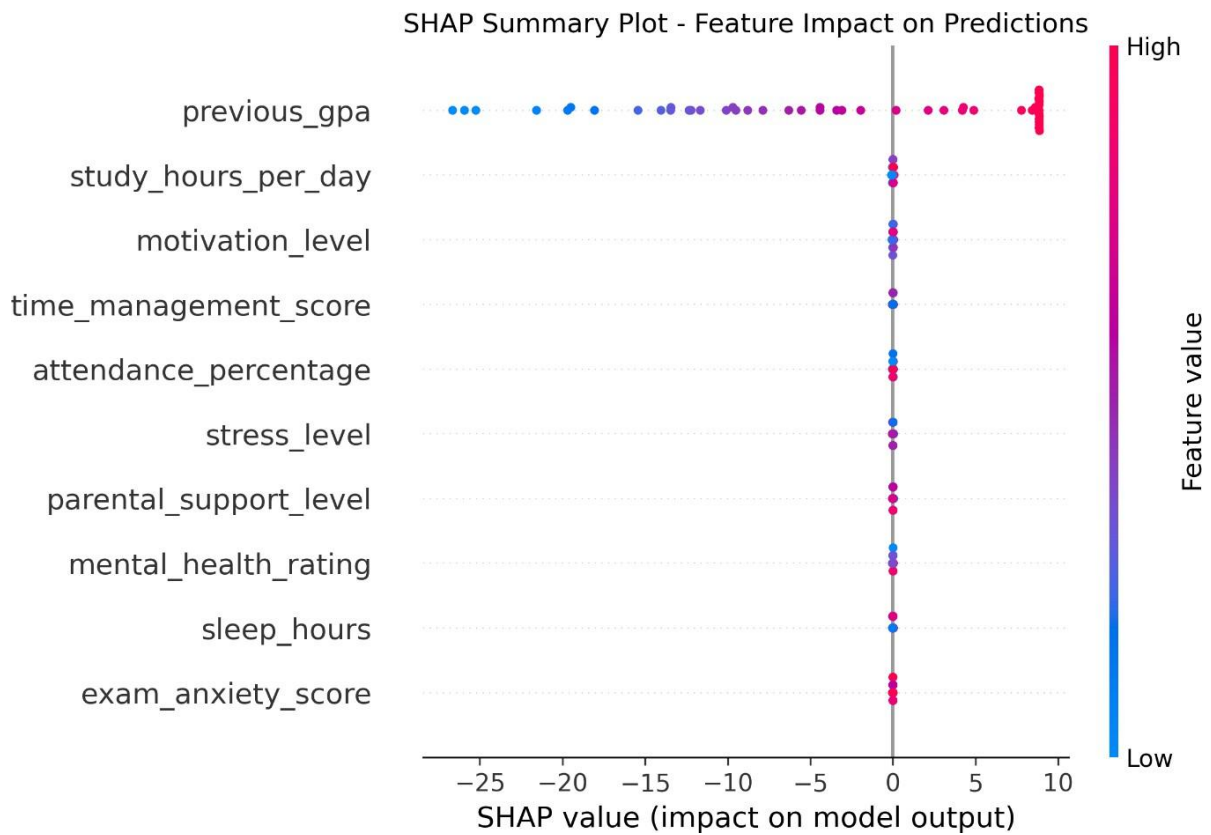


Fig.17 The SHAP summary of the stacking model

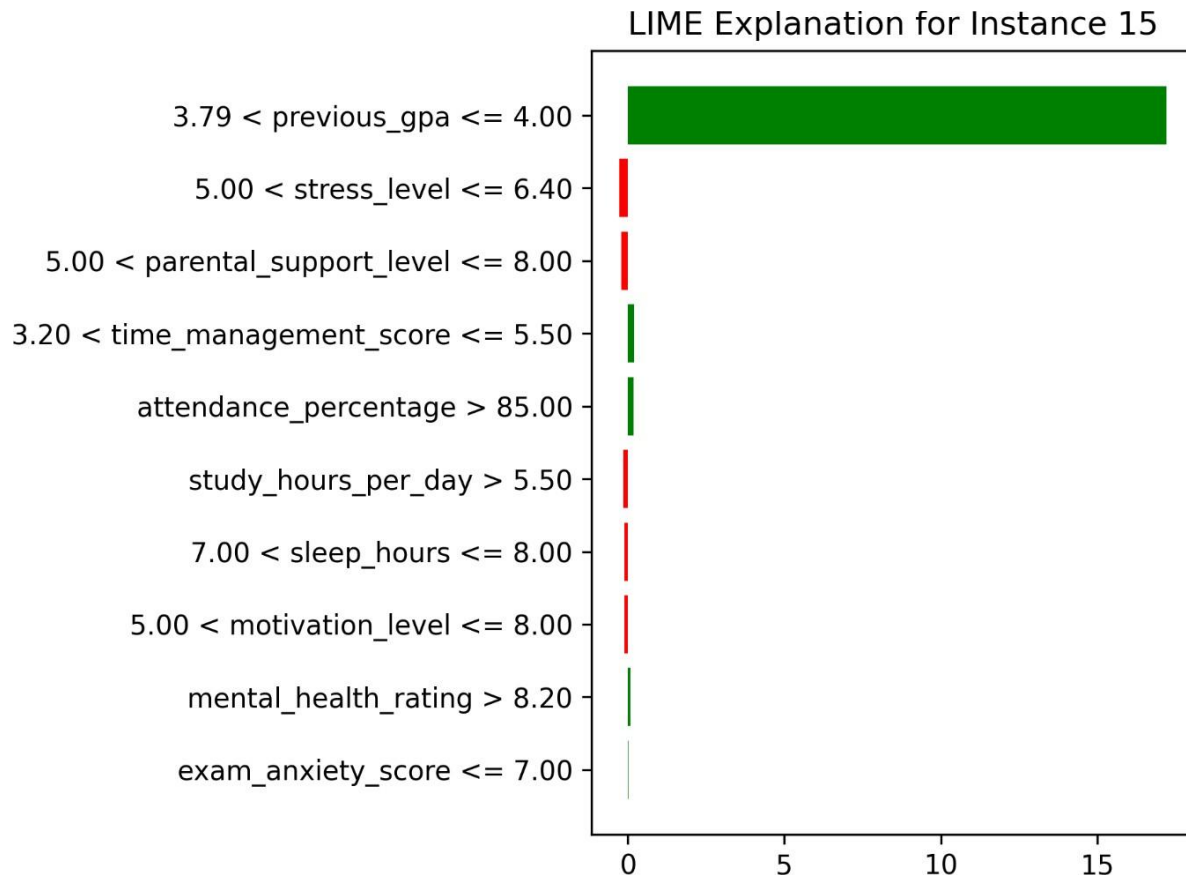


Fig.18 The LIME plot of the stacking model(instance 15)

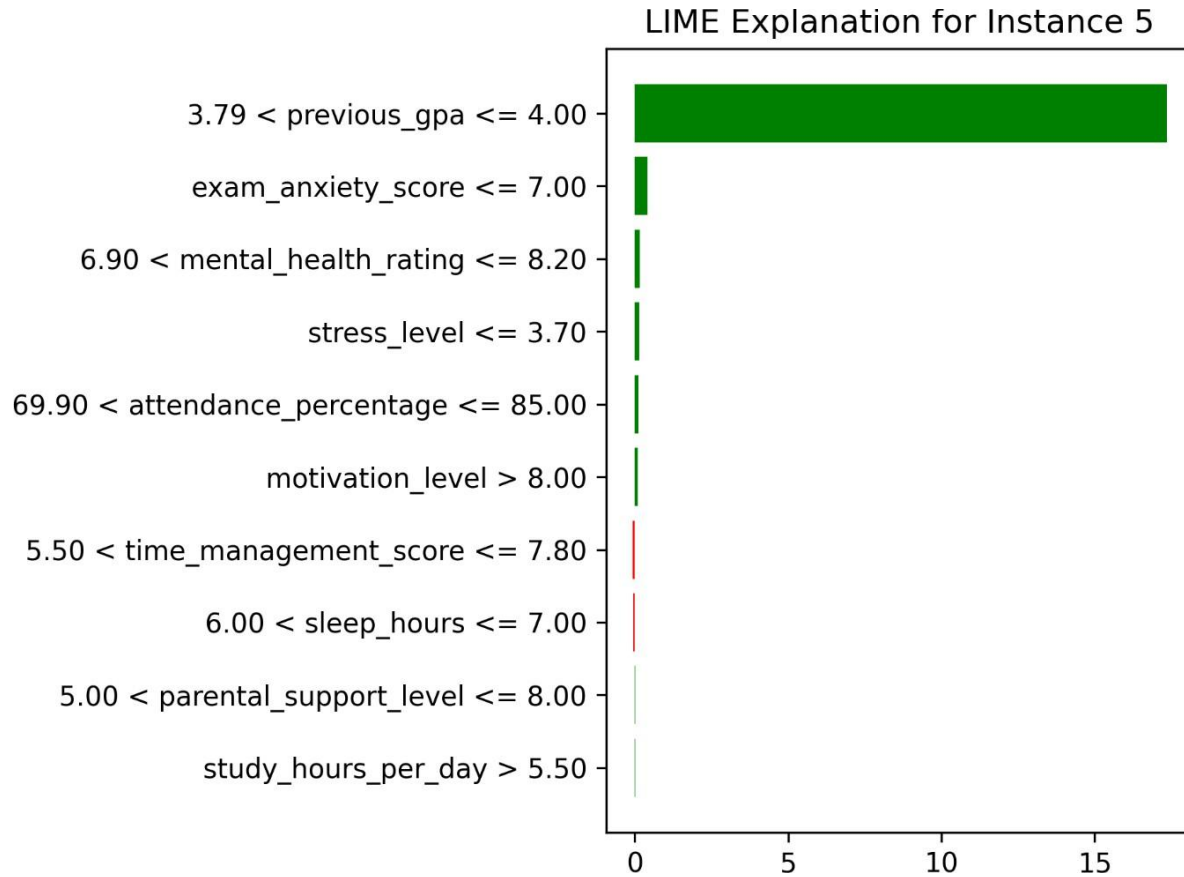


Fig.19 The LIME plot of the stacking model(instance 5)

5.4. Discussion of Findings

The results showed that the stacking model was efficient in predicting student performance with R^2 score of 0.8722 that match the Random Forest's peak, which suggesting the ensemble combines the advantages offered by different models, particularly in handling the dataset's multidimensional features like `digital_lifestyle_index` and `academic_engagement_score`. Also, the slight MSE reduction (17.3085 vs. 17.5364) showed that the MLP's role in addressing errors due to interdependent variables. We also found the KNN algorithm gave the lowest R^2 at 0.8190, which hints at challenges with local pattern and maybe overfitting because of the dataset's wide feature space. Random Forest and the Stacked Model are the

top performers, with Random Forest slightly edging out due to lower MSE and RMSE as shown in figure 15

The average impact of each feature on the model's output magnitude is depicted in Figure 16. The highest mean ($|SHAP\ value|$) is approximately 10 for `previous_gpa`, while other features, such as `exam_anxiety_score`, have negligible impacts (close to 0). The distribution of SHAP values for each feature across all instances is depicted in Figure 17. The color of the feature value (red: high, blue: low) and the position of the feature indicate the impact on the prediction (positive or negative). The prediction for instance 15 is illustrated in Figure 18, which illustrates the contributions of each feature to the output. Positive contributions are indicated by green bars, while negative contributions are indicated by red bars. For example, Figure 19 illustrates the contributions of features to its prediction, as illustrated in Figure 5. . The outcome aligns with the idea of capturing deeper interactions, but the modest impact leaves room for adjusting how the meta-learner assigns importance. Also, our framework could be more beneficial for complex datasets or a dataset that has been carefully collected and reflects the real-world habits and interactions.

5.5. Limitations of the Proposed Work

Our study faced multiple limitations that affected the scope of our study. First, while the dataset has 80,000 records with a wide set of features that make it comprehensive, it might miss some of the real-world aspects and interactions, also some of the patterns inside the data might not reflect the real-world patterns. second, a lack of long-term insights made us unable to use RNN for time-based analysis. Third, while our feature selection process was robust, it might miss finer interaction patterns because it mainly depends on statistical and wrapper techniques. Fourth, the MLP

needs a lot of hyperparameter fine-tuning and is sensitive to our hyperparameter choices. And last, the threshold-based classification lacks some accuracy.

5.6. Summary of the Chapter

This chapter detailed the performance of individual ML and DL models, with traditional models like Random Forest achieving an R^2 of 0.8722, and the stacking framework slightly enhancing this to 0.8722 with improved error metrics (MSE: 17.3085, RMSE: 4.1604). Tables and figures facilitated comparisons, revealing the ensemble's edge in leveraging diverse predictors. The discussion interpreted these gains through the lens of feature engineering and model diversity, while acknowledging limitations like synthetic data and untapped temporal analysis. These findings lay a foundation for future refinements, particularly in real-world validation and longitudinal modeling.

Chapter 6 Conclusion and Future Work

This thesis presents an innovative stacking framework for predicting student achievement, adeptly integrating machine learning (ML) and deep learning (DL) methodologies to attain superior predictive efficacy. The system improves efficiency using a hybrid "filter then wrapper" feature selection technique. The stacking ensemble combines a Multi-Layer Perceptron (MLP) meta-learner with several base learners. The stacked model has a Mean Squared Error (MSE) of 17.3085, a Root Mean Squared Error (RMSE) of 4.1604, a Mean Absolute Error (MAE) of 3.2145, and an R-squared (R^2) of 0.8722. This performance is better than the performance of each separate baseline model, showing that the framework can use the strengths of both ML and DL. To improve understanding, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were used to explain how features and models made judgments, ensuring predictions were clear. These explainability methodologies showed important factors that predict student achievement. This allowed teachers to make smart, data-driven decisions about how to help students. The framework's practical consequences include increased dropout prevention, individualized learning, and improved academic decision-making, establishing it as a significant resource for educational institutions.

Future Work:

Future research will seek to enhance the approach by:

- Integrating multi-modal data, including text-based student feedback, behavioral patterns, and temporal data from learning management systems, to identify more subtle determinants of student performance.
- Optimizing the framework for real-time prediction might make it possible to take prompt and flexible actions.
- The meta-learner should improve performance on challenging datasets by investigating more sophisticated deep learning architectures, such as transformers or recurrent neural networks.

References

- [1] Hernández-García, Á., Cuenca-Enrique, C., del Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A learning cycle approach. *Computers in Human Behavior*, 155, 108183.
- [2] Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015, July). Predicting students' performance in educational data mining. In *2015 International Symposium on Educational Technology (ISET)* (pp. 125–128). IEEE.
- [3] Isljamović, S., & Suknović, M. (2014). Predicting students' academic performance using artificial neural network: A case study from Faculty of Organizational Sciences. *The Eurasia Proceedings of Educational and Social Sciences*, 1, 68–72.
- [4] Alnasyan, B., Basher, M., & Alassafi, M. (2024). The power of deep learning techniques for predicting student performance in virtual learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100231.
- [5] Alshabandar, R., Hussain, A., Keight, R., Laws, A., Baker, T., & Radi, N. (2022). A flexible feature selection approach for predicting students' academic performance in online learning. *Computers and Electrical Engineering*, 99, 107761.
- [6] Abid, M., Elsharawy, A., & Fahmy, H. (2025). Explainable AI methods for predicting student grades and improving academic success. *Journal of Information Systems and Education Management*, 10(2), 45–59.
- [7] Zhu, H. (2020). Big data and artificial intelligence modeling for drug discovery. *Annual Review of Pharmacology and Toxicology*, 60(1), 573–589.
- [8] Rahmani, A. M., Yousefpoor, E., Yousefpoor, M. S., Mehmood, Z., Haider, A., Hosseinzadeh, M., & Ali Naqvi, R. (2021). Machine learning in medicine: Review, applications, and challenges. *Mathematics*, 9(22), 2970.

- [9] Mosavi, A., Ardabili, S., & Várkonyi-Kóczy, A. R. (2019, September). List of deep learning models. In *International Conference on Global Research and Education* (pp. 202–214). Springer.
- [10] Johnsen, M. (2024). *Large language models (LLMs)*. Maria Johnsen.
- [11] Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., ... & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9), 1–33.
- [12] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*, 50(6), 1–45.
- [13] Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. arXiv preprint arXiv:2404.16958.
- [14] Wang, Y., Ding, A., Guan, K., Wu, S., & Du, Y. (2021). Graph-based Ensemble Machine Learning for Student Performance Prediction. *arXiv preprint arXiv:2112.07893*.
- [15] Orji, F. A., & Vassileva, J. (2022). Machine learning approach for predicting students academic performance and study strategies based on their motivation. *arXiv preprint arXiv:2210.08186*.
- [16] Sandeepa, A. G. R., & Mohottala, S. (2025, February). Evaluation of machine learning models in student academic performance prediction. In *2025 5th International Conference on Advanced Research in Computing (ICARC)* (pp. 1–6). IEEE.
- [17] Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1), 4067721.
- [18] Yuan, J., Qiu, X., Wu, J., Guo, J., Li, W., & Wang, Y. G. (2024). Integrating behavior analysis with machine learning to predict online learning performance: A scientometric review and empirical study. *arXiv preprint arXiv:2406.11847*.

- [19] Arya, M., Motwani, A., Prasad, K., Dewangan, B. K., Choudhury, T., & Chauhan, P. (2024). A CNN–LSTM-based deep learning model for early prediction of students’ performance. *International Journal on Smart Sensing and Intelligent Systems*, (1).
- [20] Adefemi, K. O., Mutanga, M. B., & Jugoo, V. (2025). Hybrid deep learning models for predicting student academic performance. *Mathematical and Computational Applications*, 30(3), 59.
- [21] Ben Said, A., Abdel-Salam, A. S. G., & Hazaa, K. A. (2024). Performance prediction in online academic course: A deep learning approach with time series imaging. *Multimedia Tools and Applications*, 83(18), 55427–55445.
- [22] Junejo, N. U. R., Nawaz, M. W., Huang, Q., Dong, X., Wang, C., & Zheng, G. (2025). Accurate multi-category student performance forecasting at early stages of online education using neural networks. *Scientific Reports*, 15(1), 16251.
- [23] Ahmed, W., Wani, M. A., Plawiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15(1), 26879.
- [24] Ying, D., & Ma, J. (2024). Student Performance Prediction with Regression Approach and Data Generation. *Applied Sciences*, 14(3), 1148.
- [25] Tang, B., Li, S., & Zhao, C. (2024). Predicting the performance of students using deep ensemble learning. *Journal of Intelligence*, 12(12), 124.
- [26] Yang, K. Predicting Student Performance Using Artificial Neural Networks.
- [37] <https://www.kaggle.com/datasets/aryan208/student-habits-and-academic-performance-dataset> Last accessed:3/9/2025