



MSc in Data & Artificial Intelligence

Research Methods & Professional Development - ACFA7004_BIRM

‘Component One - Proposal’

“AI-Based Student Performance Prediction System Using ML”

Submitted To

Vijeyananthan Thayananthan
Lecturer, UWTSD Birmingham

Submitted By

Tareque Hasnat Tuhin
Mail: 2513674@student.uwtsd.ac.uk

“Submission Date”

11 November, 2025

Index

<i>SL</i>	<i>Title</i>	<i>Page</i>
#	Abstract	01
01	Introduction	03-04
1.1	Background	03
1.2	Current Situation	03
1.3	Research Problem	03
1.4	Research Aim and Objectives	04
1.5	Importance of the Research	04
02	Literature Review	04-05
2.1	Overview	04
2.2	Factors Affecting Student Performance	05
2.3	Commonly Used ML Models	05
2.4	Data Preparation and Evaluation	05
2.5	Ethical Issues	05
2.6	Research Gap	05
03	Research Design and Methodology	06
3.1	Research Design	06
3.2	Method	06
3.3	Tools and Resources	06
3.4	Ethical Considerations	06
04	Expected Results	06
05	Project Timeline	07
06	Presentation	07
07	References	07-08

Abstract

In this research, I plan to build an AI-based system that predicts student performance using machine learning. Many universities face problems identifying weak students early, which causes high failure and dropout rates. My goal is to create a simple and reliable model using student data such as attendance, grades, and class participation. I will use algorithms like decision trees, random forests, and logistic regression to test which gives better accuracy. The system will help teachers take action before students fail. I will also make sure that the model is fair, transparent, and respects data privacy. This project will show how AI can support quality education and student success while following ethical research standards.

1. Introduction

1.1 Background

Education is a key part of personal and social development. Today, technology plays a huge role in learning, and we collect more data about students than ever before: attendance, marks, and online activity. I believe this data can help teachers understand student performance better.

By using AI and ML, I aim to create a system that can predict how well students will perform before final results come out. This system can give early warnings to teachers, so they can guide and support students who might be struggling [1][2].

1.2 Current Situation

Most universities and schools still use traditional methods such as exams and manual teacher feedback to evaluate students. These are useful but slow. Often, by the time teachers realize a student is underperforming, it is too late.

AI can make this process smarter by analyzing patterns in past data to predict future performance [3]. Many institutions already use online learning platforms, so data is available but not often used effectively. My research focuses on turning that data into insight [4][5].

1.3 Research Problem

There is no simple system that can predict academic performance early and accurately. Some existing models are complex or not user-friendly for teachers [6]. Also, some models may have bias, especially when the dataset is small or unbalanced [7].

My goal is to develop a model that is accurate, fair, and easy to understand so teachers can use it confidently to support students.

1.4 Research Aim and Objectives

Aim:

To design and test a machine learning-based model that predicts student performance in higher education.

Objectives:

1. To explore existing studies on AI and ML in education.
2. To collect and prepare student academic data.
3. To train and test machine learning models such as decision trees, random forests, and logistic regression.
4. To compare model accuracy and performance using evaluation metrics.
5. To ensure fairness and reduce bias in the predictions.
6. To design an easy-to-understand output for teachers.
7. To maintain ethical standards and protect data privacy.

1.5 Importance of the Research

This research is important because it can help universities and teachers identify weak students early and provide timely help. It will also reduce failure rates and support better student outcomes [8].

AI tools like this can improve teaching quality, make assessments fairer, and contribute to the United Nations goal of quality education for all [9][10].

2. Literature Review

2.1 Overview

I reviewed several research studies about AI in education. Many focused on predicting student grades and dropout rates using machine learning. Most of these studies used academic data like attendance, previous grades, and online activity [11][12].

However, I noticed that some models are not very transparent, and teachers find it difficult to interpret the predictions. My goal is to build a model that is both accurate and explainable [13].

2.2 Factors Affecting Student Performance

Different factors can affect how a student performs: attendance, class participation, motivation, social background, and study hours [14]. For this project, I will use mainly attendance, assignment marks, and exam scores, as they are measurable and reliable [15].

2.3 Commonly Used ML Models

Studies have shown that models like decision trees, random forests, naïve Bayes, and support vector machines perform well for educational data [16]. Decision tree models are easy to understand, while random forest usually gives better accuracy [17]. Logistic regression is simple but powerful for binary predictions (pass/fail) [18].

2.4 Data Preparation and Evaluation

Data cleaning and preparation are very important. Missing or wrong values can make the model inaccurate [19]. I will normalize and split the data into training and testing sets. Model performance will be evaluated using accuracy, precision, recall, and F1-score [20].

2.5 Ethical Issues

Since I am working with educational data, I will make sure all information is anonymous. I will follow GDPR guidelines and store data safely. I will also make sure that my model does not unfairly target any group of students [21][22].

2.6 Research Gap

Many past studies focused on specific datasets and didn't discuss fairness or explainability. I will address this gap by ensuring my system is understandable for teachers and fair across all groups.

3. Research Design and Methodology

3.1 Research Design

I will follow a design science research approach. This means I will identify a problem, design an artifact (ML model), test it, and refine it based on results. This is suitable because I want to create something practical that others can use.

3.2 Method

This research is quantitative, as I will work with numerical data.

Steps I will follow:

1. Collect anonymized student data.
2. Clean and preprocess the dataset.
3. Train machine learning algorithms.
4. Compare model performance.
5. Analyze the results and visualize predictions.

3.3 Tools and Resources

- ★ Software: Python, Scikit-learn, Pandas, Matplotlib.
- ★ Platform: Google Colab.
- ★ Hardware: Laptop or cloud GPU if needed.
- ★ Dataset: Open-source educational dataset (UCI repository).

3.4 Ethical Considerations

I will not collect personal information such as names or student IDs. I will use only anonymized data for academic purposes. I will store all data securely and delete it after completing the project. I will also discuss the research with my supervisor to ensure full ethical approval.

4. Expected Results

I expect the Random Forest model to give the best results in accuracy and reliability. The output will show predicted performance levels, allowing teachers to take early action.

The study may also reveal which factors affect performance most strongly.

Possible limitations:

- Limited access to real data.
- Bias in small datasets.
- Time required for data cleaning.

5. Project Timeline

Tasks	Jan	Feb	Mar	Apr
Topic Selection				
Literature Review				
Data Collection				
Model Development				
Testing & Evaluation				
Writing & Editing				
Final Submission				

6. Presentation

I will write the final report in simple English with clear headings and short paragraphs. I will include graphs, tables, and charts for clarity. The report will follow a logical order from background to conclusion.

7. References

- [1] T. Hossain, "AI in Modern Education," Learning Tech Journal, 2021.
- [2] M. Alam, "Machine Learning for Academic Success," Education Today, 2022.
- [3] L. Rahman, "Predicting Student Grades with AI," Open Learning Review, 2020.
- [4] J. Smith, "Data-Driven Education," Global EdTech Reports, 2023.

- [5] A. Khan, "Improving Education with Predictive Analytics," Computing Education Journal, 2021.
- [6] N. Das, "AI Applications in Higher Learning," Tech in Education, 2024.
- [7] S. Roy, "Understanding Academic Data Patterns," Student Success Review, 2019.
- [8] P. Green, "Challenges in Educational Data Mining," Smart Learning Press, 2018.
- [9] D. Karim, "AI Tools in Schools," Digital Learning, 2023.
- [10] R. Akter, "Predictive Models in Education," Bangladesh Computing Journal, 2022.
- [11] M. Chowdhury, "The Future of AI in Education," Smart EduTech Journal, 2021.
- [12] J. Lopez, "A Review on Student Performance Prediction," Tech Learning Review, 2017.
- [13] A. Nahar, "Early Warning Systems for Students," Open Research Education, 2020.
- [14] F. Rahim, "Key Factors Affecting Academic Results," Education Studies Review, 2019.
- [15] G. Ahmed, "The Impact of Attendance on Student Success," Higher Studies Review, 2018.
- [16] L. Islam, "Using ML Models to Predict Results," Journal of AI Applications, 2021.
- [17] S. Miah, "Decision Tree for Student Analytics," Data and Learning Journal, 2020.
- [18] A. Jahan, "Neural Networks in Education Prediction," AI Computing Review, 2023.
- [19] T. Hasan, "Data Cleaning in Academic Research," Research Data Notes, 2022.
- [20] UK Information Commissioner's Office, "GDPR for Researchers," 2023.
- [21] K. Ahmed, "Bias in AI Systems," Ethics and Technology Review, 2023.
- [22] R. Uddin, "Fair and Explainable AI Models," Modern Computing Research, 2024.