**MMDT_2024_094**

**Project Title: Early Prediction of Student Final Exam Results**

**MemberID:** mmdt2024.094

**Mentor:** Ma Khin Su Myat

**Scenario**

Myanmar's education system has faced many challenges in recent years, especially due to ongoing political instability and social unrest. School closures, teacher shortages, and disruptions have affected the quality of learning for many students. Amidst this difficult environment, many parents are still very eager to help their children succeed by encouraging them to take international exams like the IGCSE, hoping it will open doors for studying abroad.

Firstly, I want to create a system that **recommends whether a student should sit for the ICT Global Exam or not** based on their past performance, like attendance, homework, and midterm scores. This recommendation will help students and teachers decide if the student is ready for the exam. I have seen how parents are deeply committed to giving their children the best possible education. Many of them save money for months just to afford IGCSE exam fees. But with this investment comes anxiety:

**"Is my child truly ready to take the exam, or are we rushing into it too soon?"**

But I chose this project the project title "Early Prediction of Student Final Exam Results" because starting directly with exam recommendation system is that felt difficult for me. So, I decided to first build a prediction final score system to guide students which is a good first step for me.

## I. Project Description:

This project is to **predict how well students will do on their final exam** using information like their attendance, homework scores, and midterm exam results. By making these predictions early, teachers can find out which students might struggle and help them before the final exam.

1. **Why Did You Choose This Problem?**

   I chose this problem because it is important to help students succeed. Sometimes teachers don't know which students need more help until it's too late. By predicting final exam scores early, teachers can support students better and improve their chances of passing.

2. My goal is to predict students' final exam scores based on their attendance, homework scores, and midterm results.

## II. Dataset

The dataset is **synthetic datasets that** based on the features used in my project.

**What features are used?**

| Features | Description |
|---|---|
| Mock Exam Score | How well the student did in practice exams. |
| Practical Score | How the student performed in hands-on tasks. |
| Theory Score | The student's marks in written tests. |
| Attendance | Percentage of classes the student attended. |
| Revision Time | Hours spent studying outside class. |
| Final Exam Score: | The actual score in the final exam (**target variable**). |

**Why these features?**

- **Mock Exam Score:** How well the student did in practice exams.
- **Practical Score:** How the student performed in hands-on tasks.
- **Theory Score:** The student's marks in written tests.
- **Attendance:** Percentage of classes the student attended.
- **Average Score:** The student's overall average from different tests.
- **Final Exam Score:** The actual score in the final exam (what we want to predict).

## III. Modelling Approach

In this project, it is used a Polynomial Regression model with a Linear Regression to predict the **final exam score.** Linear regression is simple, easy to understand, and works well. built this model using Linear Regression from scikit-learn, but applied Polynomial Features to capture non-linear relationships between input features (like revision time, attendance, etc.) and the final exam score. To find the best polynomial degree and used **GridSearchCV**, which automatically tested multiple degrees (2, 3, 4, 5, 7, 9) and picked the one that performed best.

By project reference, I also applied a **Pipeline**, which made the code clean and organized and also used:

a. **StandardScaler** – scaled all input features so they have **mean = 0** and **standard deviation = 1**.
b. **Polynomial Features**: This allows the model to capture more complex relationships between features and the target.
c. **Linear Regression** – Trained the model on the transformed data to learn how different factors influence the final exam score.

**I choose these** approaches:

- It is simple but powerful: **Linear Regression and Polynomial Features** can model both straight-line and curved patterns. And it is also automated because Grid search found the **best polynomial degree** using cross-validation.

## IV. Model Evaluation

To assess how well the model predicted students' final exam scores, I used three key evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and the $R^2$ Score.

I selected a Polynomial Regression model with degree = 2, which allowed the model to capture some non-linear patterns in the data while keeping it simple and efficient. By using a polynomial regression model of degree 2, I achieved excellent results. The model accurately predicted student performance with almost no error and explained nearly all the variation in exam results. This suggests that the chosen features (like mock scores, attendance, and revision time) strongly influence final outcomes and were modeled effectively.

**Model Evaluation Results (Polynomial Degree = 2)**

| Metric | Value | Explanation |
|---|---|---|
| Mean Absolute Error | 0.0 | The model's predictions were nearly perfect — no average error. |
| Mean Squared Error | 7.37 | Very small squared error, showing very few and very minor mistakes. |
| $R^2$ Score | 0.99 | The model explained almost 100% of the variation in final exam scores. |

## V. Reflection

At first, I found it difficult to choose a good project idea. I originally wanted to build a system to **recommend whether students should sit the IGCSE exam**, but it felt too complex for me as a beginner. So, I simplified the goal to predicting students' final exam scores. Another difficulty was the lack of real-world data. Since I did not have access to actual student records, I had to create a realistic synthetic dataset myself, which involved understanding how to design features and clean the data properly. Throughout this project, I learned a lot about the machine learning workflow. I became more confident in using Python libraries such as Pandas, Scikit-learn, and GridSearchCV. I learned how to select relevant features for the model, split data into training and testing sets, and apply scaling techniques like StandardScaler. One of the most valuable lessons was understanding how polynomial regression can model non-linear relationships in the data. I also learned how to evaluate model performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and the $R^2$ score. These helped me assess how accurate and reliable the model was in predicting student scores. In the future, I would like to expand this project to not only predict student exam scores but also give a recommendation. For example, whether a student is ready to sit the IGCSE ICT exam or should spend more time preparing**.** This would be based on the predicted score and a set threshold (e.g., if predicted score ≥ 60%, recommend to sit; if below, recommend more preparation). It would help both teachers and students make smarter decisions. Finally, I would create a simple app or dashboard using Streamlit so that teachers can easily use the model without writing any code.