CEN 420 01 Introduction to Pattern Recognition
Project Proposal

Meaad Farag Bayuosef

***Abstract*—This document presents a proposal for the project "Student Performance Predictor" conducted by the students of the Software Engineering Department at Maltepe University.**

I. Introduction

This project aims to build a Student Performance Predictor using machine learning techniques. By predicting students' academic performance based on various factors, this project has practical applications in education.

II. Problem Description

This project seeks to develop a machine learning model to predict student performance based on demographics, study habits, attendance, and past academic records. Identifying at-risk students allows educational institutions to provide targeted interventions for academic success.

III. Project Scope

A. Topic Selection

The chosen topic, "Student Performance Predictor," involves predicting students' academic performance based on historical data and relevant features.

B. Data Collection

Gather data related to students' academic records, including variables such as attendance, previous grades. This data will be crucial for training and evaluating the machine learning model.

C. Algorithm Selection

Identify and implement a suitable machine learning algorithm for regression tasks. Supervised Learning could be a potential candidate to predict continuous variables such as grades.

D. Model Evaluation

Implement cross-validation techniques to ensure the model's generalization ability. The model's performance will be evaluated based on metrics like accuracy, precision, recall, F1-score, Mean Squared Error, or R-squared.

IV. Methodology

A. Data Preprocessing

Clean and preprocess the data by handling missing values, encoding categorical variables, and normalizing numerical features.

B. Feature Selection

Identify the most influential features affecting academic performance through exploratory data analysis and statistical methods.

C. Model Development

Implement and train the chosen machine learning algorithm using historical data.

D. Model Evaluation

Evaluate the model's performance using appropriate metrics and fine-tune the model if necessary.

V. DATASET DETAILS

- Size: 5,000 student records
- Features: Demographics (age, gender, etc.), study hours, attendance, parental education, past exam scores, etc.
- Data Quality: Mostly complete, minimal missing values, and outliers treated.

VI. PLANNED PERFORMANCE EVALUATION METHODS

We will evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. Additionally, we will establish baseline performance using traditional statistical methods such as regression analysis.

VII. EXPECTED OUTCOMES

- A trained machine learning model capable of predicting student performance accurately.
- Insights into the significant factors influencing academic success.

VIII. Project Timeline

- Weeks 1-3: Project selection and data collection.
- Weeks 4-5: Data Cleaning and Preprocessing.
- Weeks 6-7: Feature Selection and Engineering.
- Weeks 8-9: Algorithm Implementation, Model Training.
- Weeks 10-11: Model Evaluation and Fine-Tuning.
- Weeks 12-13: Report Writing and Documentation.
- Week 14-15: Project Submission and Presentation.

IX. Conclusion

The Student Performance Predictor project applies machine learning in education, offering insights into factors affecting academic success and improving educational outcomes.