IDEA PROPOSAL

Group members: Mahsa Hamidi, Noorullah Zamindar, Ahmad Reshad Amir beag, Abdulrawof Totakhil, Eshaq Karimi, Yaser Zarifi

Project Idea:

Our project aims to predict students' math scores based on various features such as gender, parental level of education, race, lunch type, test preparation course, writing scores, and reading scores. By applying supervised machine learning techniques, we seek to identify key factors that impact student performance in mathematics, thereby providing insights to educators and policymakers.

Relevance to Sustainable Development Goals (SDGs):

This project aligns with **SDG4**: Quality Education. By understanding the factors that influence students' math performance, we can develop targeted interventions to improve educational outcomes. This contributes to ensuring inclusive and equitable quality education and promotes lifelong learning opportunities for all.

Literature Examples:

- A study of factors affecting students' performance in examination at university level: This paper explores various determinants of student performance and provides a comparative analysis of different influencing factors. It offers valuable insights into understanding student performance metrics and can serve as a benchmark for our project.
- Literature Review of Analyzing and Predicting Students' Performance in Examinations: This review paper discusses various methodologies and models used for predicting student performance, highlighting both their strengths and limitations. It serves as a foundational reference for selecting appropriate machine learning techniques in our study.

Describe Your Data:

The dataset, titled "Student Performance in Exams," consists of the following features: gender, parental level of education, race, lunch type, test preparation course, math score, writing score, and reading score. The data is in CSV format and can be accessed from here. Preprocessing steps include handling missing values, encoding categorical variables, and normalizing numeric features.

Approach (Machine Learning or Deep Learning):

We will use supervised machine learning techniques such as Linear Regression, K-Neighbors Regressor, Decision Tree, Random Forest Regressor, XGBoost Regressor, CatBoost Regressor, and AdaBoost Regressor to predict math scores. Given the structure and size of our data, a machine learning approach is suitable due to its efficiency and effectiveness in handling tabular data with both numeric and categorical features.