

PROJECT REPORT

Project Name: Student Performance Prediction Model

1. Problem Analysis

The project addresses the problem of predicting a student's final grade in a college course. This is a multi-class classification problem, where the final grade is categorized into distinct classes (e.g., A, B, C, D, F). The goal is to provide insights into factors influencing academic success and enable early identification of students who may require additional support. The prediction is based on a dataset containing student information, including:

- 'GPA': High School GPA
- 'Extracurriculars': Number of extracurricular activities
- 'Parental_Income': Parental income
- 'Study_Hours': Number of hours studied per week
- 'Entrance_Score': Performance in an entrance test
- 'Distance': Distance from college
- 'Internet': Internet availability at home (Yes/No)

The target variable is 'Final_Grade'.

2. Algorithm Selection

The Decision Tree Classifier algorithm was chosen for this project. This algorithm is well-suited for multi-class classification problems and can handle both numerical and categorical data, as present in the dataset. Decision Trees offer a balance of predictive performance and interpretability. The tree structure provides a clear representation of the decision-making process, showing which factors (features) are most influential in predicting student grades.

3. Model Overview

The model development process involved the following steps:

- **Data Loading:** The dataset ("merged_college_data.csv") was loaded using Pandas.
- **Data Preprocessing:**
 - Numerical features were standardized using StandardScaler.
 - The categorical feature ('Internet') was one-hot encoded using OneHotEncoder.
 - ColumnTransformer was used to apply these preprocessing steps.
- **Model Training:** A Decision Tree Classifier was trained on the preprocessed data. The data was split into training (80%) and testing (20%) sets, ensuring class balance using stratified sampling.
- **Model Evaluation:** The model's performance was evaluated on the test set using accuracy, a classification report, and a confusion matrix.
- **Prediction on New Data:** The predict_grade() function was defined to predict grades for new students.

The Decision Tree Classifier provides a reasonable approach to predicting student final grades, leveraging a combination of numerical and categorical features.