

Project Documentation

Year, Branch, Section: 2nd Year, Computer Science and Data Science

Team Number: 2

Team Lead Name: Sri Naga Durga

Team Members Details:

- Kuncham Durga Prasad: 23B21A4420
- P. Venu: 23B21A4419
- V. Lasya Gowri Pravallika: 23B21A4404
- P. Divya: 23B21A4407

Problem Statement and Small Matter About It:

Problem Statement: Predicting Student Grades.

Matter: In the realm of education, proactive intervention is paramount to student success.

Predicting student grades allows for the identification of at-risk students, enabling educators to provide targeted support and resources. This project aims to develop a predictive model using machine learning to forecast student academic performance, thereby facilitating early intervention and improving overall educational outcomes.

Domain: Predicting Student Grades (Educational Data Mining)

Technologies Used:

- **Programming Language:** Python
- **Libraries:**
 - Pandas (for data manipulation)
 - NumPy (for numerical operations)
 - Scikit-learn (for machine learning models and evaluation)
 - Matplotlib and Seaborn (for data visualization)
- **Machine Learning Algorithms:**
 - Linear Regression
 - Random Forest Regression
 - Decision Tree Regression
 - Support Vector Regression
- **Data Storage:** CSV files

Approach to Do Project:

1. **Data Collection:**
 - Acquired student data from available sources, including academic records (grades, attendance, assignment scores), and potentially relevant demographic information.
2. **Data Preprocessing:**
 - Cleaned the dataset by handling missing values and outliers.
 - Transformed categorical features into numerical representations (e.g., one-hot encoding).
 - Normalized or standardized numerical features.
 - Feature engineering and selection was performed to improve model performance.
3. **Exploratory Data Analysis (EDA):**
 - Visualized data distributions and relationships using plots and charts.
 - Calculated descriptive statistics to understand data characteristics.
 - Identified correlations between features and the target variable (grades).
4. **Model Training:**

- Split the dataset into training and testing sets.
- Trained multiple regression models on the training data.
- Hyperparameter tuning using cross validation.

5. Model Evaluation:

- Evaluated model performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.
- Compared the performance of different models to select the best one.

6. Reporting:

- Documented the entire project process, including methodology, results, and conclusions.

Outcome:

- A trained machine learning model capable of predicting student grades.
- An analysis of the factors influencing student academic performance.
- A comprehensive report summarizing the project.

Conclusion:

This project successfully applied machine learning techniques to predict student grades. The developed model demonstrates the feasibility of using data-driven approaches to enhance educational practices. By identifying key factors that impact student performance, institutions can implement targeted interventions to improve academic outcomes. This project highlights the potential of data science in contributing to a more effective and personalized learning environment.