

Documentation and observations:

The project aims to predict the CGPA of a student based on different grades in four years. The dataset used for this project is called "Grades.csv" and can be found on GitHub[1]. The dataset contains grades scored by students throughout their university tenure in various courses and their CGPA calculated based on their grades. The dataset has 43 columns, including the "Seat No" column which is the enrolled number of candidates that took the exams, and the "CGPA" column which is the cumulative GPA based on the four-year total grade progress of each candidate. The other columns are course codes in the format AB-XXX where AB are alphabets representing candidates' departments and XXX are numbers where the first X represents the year the candidate took the exam.

The first step in the project is to import the necessary libraries, including pandas, numpy, sklearn, matplotlib, and seaborn. Then, the dataset is read into a pandas dataframe using the `pd.read_csv()` function. The dataset has 571 rows and 43 columns. The "Seat No" column is dropped from the dataframe as it is not needed for the prediction.

The dataset is then explored by looking at the first few rows using the `head()` function, and the basic statistics of the CGPA column are calculated using the `describe()` function. The mean CGPA is 2.95, with a standard deviation of 0.62, and a minimum CGPA of 0.8 and a maximum CGPA of 3.98.

Next, the data is prepared for training the model by splitting the data into training and testing sets using the `train_test_split()` function from sklearn. The features and target variables are defined, and the data is scaled using the `StandardScaler()` function from sklearn.

Finally, a linear regression model is trained using the `LinearRegression()` function from sklearn, and the model's performance is evaluated using the mean squared error (MSE) and the coefficient of determination (R-squared) metrics.

Overall, the project demonstrates a basic approach to predicting student grades using machine learning techniques. The project could be further improved by exploring other machine learning algorithms, feature engineering, and hyperparameter tuning.

## what machine learning techniques are used:

The `student_grades_prediction_project_6.ipynb` file uses linear regression, a type of machine learning algorithm, to predict the cumulative grade point average (CGPA) of a student based on their grades in various courses throughout their university tenure. The project involves importing

necessary libraries, reading the dataset from GitHub, preparing the data for training, and training a linear regression model to make predictions. The dataset contains grades scored by students in different courses, and the CGPA is calculated based on their four-year total grade progress. The project aims to use machine learning techniques to predict the CGPA of a student based on their grades in different courses.

#### Conclusions:

The student grades prediction project in the `student\_grades\_prediction\_project\_6.ipynb` file uses a linear regression model to predict the cumulative grade point average (CGPA) of a student based on their grades in various courses throughout their university tenure. The model is trained using the `LinearRegression()` function from the scikit-learn library. The model's performance is evaluated using the mean squared error (MSE) and the coefficient of determination (R-squared) metrics.