**Title Of Project**: Predict Marks of a faction of student of a given Data set using Machine Learning Models

**Student Name :** Adityya Jha

**Enrollment Number:** 12219011921

Batch :AIDS B2-B

G-Drive Link: <https://colab.research.google.com/drive/1My22sGpcKE5xbXVh2S9zCexCInRrMX81>

Discord Link:

<https://github.com/adityaJhaa/ML_model_for_marks_prediction>

Abstract:In this project, we explored the use of machine learning algorithms for regression analysis. The dataset consisted of student grades, with various features such as intermediate exam scores and final exam grades. The objective was to build regression models to predict the total grade based on these features.

First, the dataset was loaded and preprocessed using the pandas library. The features and target variable were separated into input (X) and output (y) variables. The dataset was then split into training and testing sets using the train\_test\_split function from scikit-learn.

Three regression models were implemented and evaluated: Linear Regression, Decision Tree Regressor, and Random Forest Regressor. For each model, the training data was used to fit the model, and the test data was used for prediction. Various evaluation metrics were employed to assess the models' performance, including mean squared error (MSE), mean absolute error (MAE), and R-squared.

The Linear Regression model provided accurate predictions, as evidenced by an R-squared score close to 1 and low values for MSE and MAE. The Decision Tree Regressor and Random Forest Regressor models also exhibited satisfactory performance, although their results varied slightly from the Linear Regression model.

These regression models can be further fine-tuned and optimized to enhance their predictive capabilities. The project demonstrated the effectiveness of using machine learning algorithms for regression tasks, highlighting the importance of feature selection, model selection, and evaluation metrics to achieve accurate predictions.

Overall, this project showcased the application of regression analysis and machine learning techniques in predicting student grades based on various factors. The models developed in this project can be useful in educational institutions for understanding student performance and identifying factors that influence grades.

**Proposed Methodology:**

1. Dataset Details: The dataset used in this project consists of student grades. It contains the following columns:

* **Intermediate exam scores**: These columns represent the scores obtained by students in intermediate exams. Each column corresponds to a specific subject or exam.
* **Final exam grade**: This column represents the final grade obtained by each student.
* **Total (100 points)**: This column represents the total grade obtained by each student, which is the target variable.

1. Preprocessing: The dataset requires preprocessing steps to prepare it for modeling. The following preprocessing steps can be applied:

* **Handling missing values**: If there are any missing values in the dataset, they can be handled by imputing them with appropriate values or using techniques such as mean imputation or interpolation.
* **Feature selection**: It is important to identify relevant features for the regression task. If there are irrelevant or redundant features, they can be removed from the dataset to improve model performance.
* **Scaling**: If the features have different scales, it is recommended to apply feature scaling techniques such as standardization or normalization to ensure that all features contribute equally to the regression model.
* **Categorical variable encoding**: If there are categorical variables in the dataset, they can be encoded using techniques like one-hot encoding or label encoding, depending on the nature of the variables and the requirements of the regression model.

1. Model Evaluation and Future Work: Once the dataset is preprocessed, regression models can be built and evaluated using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared. In this project, **Linear Regression, Decision Tree Regressor, and Random Forest Regressor models** were used.

For future work, several aspects can be considered:

* **Hyperparameter tuning**: The performance of the regression models can be further improved by tuning the hyperparameters using techniques like grid search or randomized search.
* **Cross-validation**: Instead of a single train-test split, cross-validation techniques such as k-fold cross-validation can be applied to obtain more reliable and robust model evaluation results.
* **Feature engineering**: Additional features can be created based on domain knowledge or by exploring interactions and transformations of existing features. This can potentially improve the predictive power of the models.
* **Ensemble methods**: Further exploration of ensemble methods, such as combining multiple regression models, can be performed to leverage the strengths of different models and improve overall performance.
* **Model interpretation**: Techniques for interpreting the regression models, such as feature importance analysis, can be applied to gain insights into the factors influencing student grades and enhance the understandability of the models.

By following this proposed methodology, the project can be extended to enhance the regression models' performance, gain insights from the data, and provide valuable information for educational institutions to understand and predict student grades more effectively.