

# Project Report

## Title: Predicting Student Results Using Random Forest Classifier

### ABSTRACT:

This project aims to predict student results based on various performance metrics using a Random Forest Classifier. The dataset includes marks in Python, Stats and ML, and SQL, along with their respective feedback. The project involves data preprocessing, feature engineering, feature selection, model training, and evaluation. The primary objective is to build a model that accurately predicts whether a student has passed or failed based on their performance metrics.

**PROJECT DESCRIPTION:** Build a simple ML model (classification model) to predict the 'Result' of the students who appeared for the Placement Readiness Test held on 2nd June, 2024 for Data Science and AI.

### DATA DESCRIPTION:

The dataset contains the following columns:

- **Email:** Email addresses of the individuals.
- **Python Marks:** Marks obtained in Python.
- **Python %:** Percentage score in Python.
- **Stats and ML Marks:** Marks obtained in Statistics and Machine Learning.
- **Stats and ML %:** Percentage score in Statistics and Machine Learning.
- **SQL Marks:** Marks obtained in SQL.

- **SQL %:** Percentage score in SQL.
- **Result:** Pass or Fail result based on the marks.
- **Python Feedback:** Feedback on Python.
- **SQL Feedback:** Feedback on SQL.

## STEPS INCLUDED:

- Loading the dataset.
- Dropping unnecessary columns.
- Encoding the Result column.
- Removing outliers using the IQR method.
- Feature engineering by adding a new feature Total %.
- Scaling features using StandardScaler.
- Feature selection using Recursive Feature Elimination (RFE).
- Splitting the data into training and testing sets.
- Training a Random Forest Classifier.
- Evaluating the model's performance and printing the classification report.

## RESULT:

Finally, we trained multiple classification models like **Logistic Regression**, **Decision Tree classifier** and **Random Forest classifier** model on the selected features to predict whether a student's result falls into a certain category based on their performance metrics (Python %, Stats and ML %, SQL %, etc.). The prediction is **evaluated** using a **classification report** and **accuracy score** which includes metrics like **precision**, **recall**, and **F1-score**.

## CONCLUSION:

The project demonstrates the use of multiple classification models to predict student results based on their performance metrics. The process involves **data preprocessing**, **feature engineering**, **feature selection**, and **model evaluation**. The model's performance is evaluated using standard classification metrics..