Assessment Report

on

“Student Performance Prediction"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

CSE AIML(C)

By

Saumya Sharma

Roll no. 168 (22)

Under the supervision of

Mr. Abhishek Shukla

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow
(Formerly UPTU)

18 May, 2025

Introduction

Student performance prediction is a key challenge in the education domain. By analyzing attributes such as attendance, study habits, and participation in extracurricular activities, we can predict whether a student is likely to pass or fail. This helps educators and institutions take proactive steps for academic improvement.

The aim of this project is to build a classification model that predicts a student's academic outcome (Pass/Fail) using machine learning techniques.

Methodology

1. Dataset Overview

We used a dataset containing features like:

- Attendance (Absences)
- Weekly Study Time
- Participation in tutoring, extracurriculars, volunteering, sports, etc.
- Parental support and education
- Age and GPA

2. Data Preprocessing

- Created a binary target column Pass, where students with GPA ≥ 2.0 are labeled as 1 (Pass), otherwise 0 (Fail).
- Selected 10 relevant features and split the dataset into training and testing sets (80/20 split).

3. Model Used

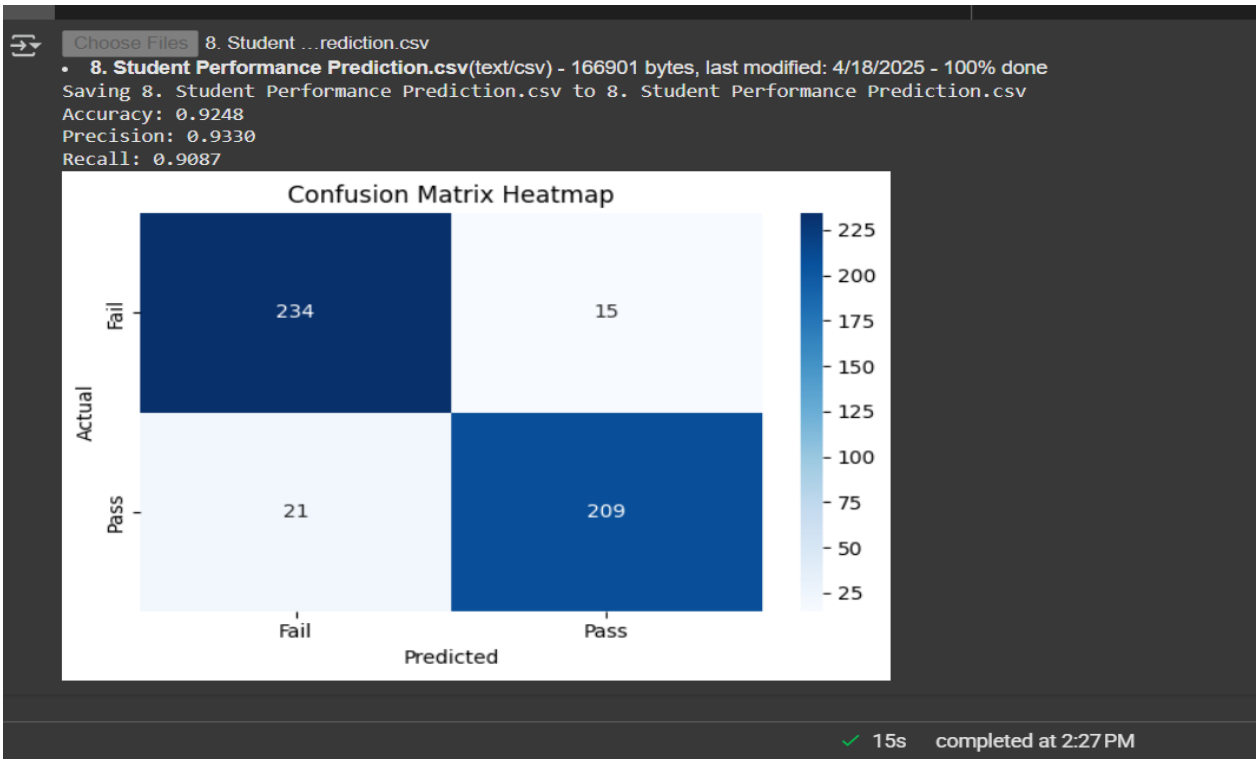
- A **Random Forest Classifier** was trained to handle the classification task.
- Evaluation was done using confusion matrix, accuracy, precision, and recall.

Code

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score

# Load the CSV
df = pd.read_csv("8. Student Performance Prediction.csv")

# Create binary target
df['Pass'] = (df['GPA'] >= 2.0).astype(int)

# Select features
features = [
    'Absences', 'StudyTimeWeekly', 'Tutoring',
    'ParentalSupport', 'Extracurricular', 'Sports',
    'Music', 'Volunteering', 'ParentalEducation', 'Age'
]
X = df[features]
y = df['Pass']

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Train model
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Evaluation
conf_mat = confusion_matrix(y_test, y_pred)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

# Print scores
print(f"Accuracy: {accuracy:.4f}")
print(f"Precision: {precision:.4f}")
print(f"Recall: {recall:.4f}")
```

```
# Heatmap
plt.figure(figsize=(6, 4))
sns.heatmap(conf_mat, annot=True, fmt='d', cmap='Blues',
        xticklabels=['Fail', 'Pass'], yticklabels=['Fail', 'Pass'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix Heatmap')
plt.tight_layout()
plt.show()
```

Result



- **Accuracy**: 92.48%
- **Precision**: 93.30%
- **Recall**: 90.87%

The confusion matrix heatmap indicates that the model performs well in classifying both Pass and Fail cases. Random Forests handled the task efficiently with minimal tuning.

References

1. Dataset Source: *[Provided by Instructor / Assignment Portal]*
2. Scikit-learn Documentation: https://scikit-learn.org
3. Seaborn & Matplotlib Documentation: https://seaborn.pydata.org | https://matplotlib.org
4. "Introduction to Machine Learning" – Andrew Ng, Coursera