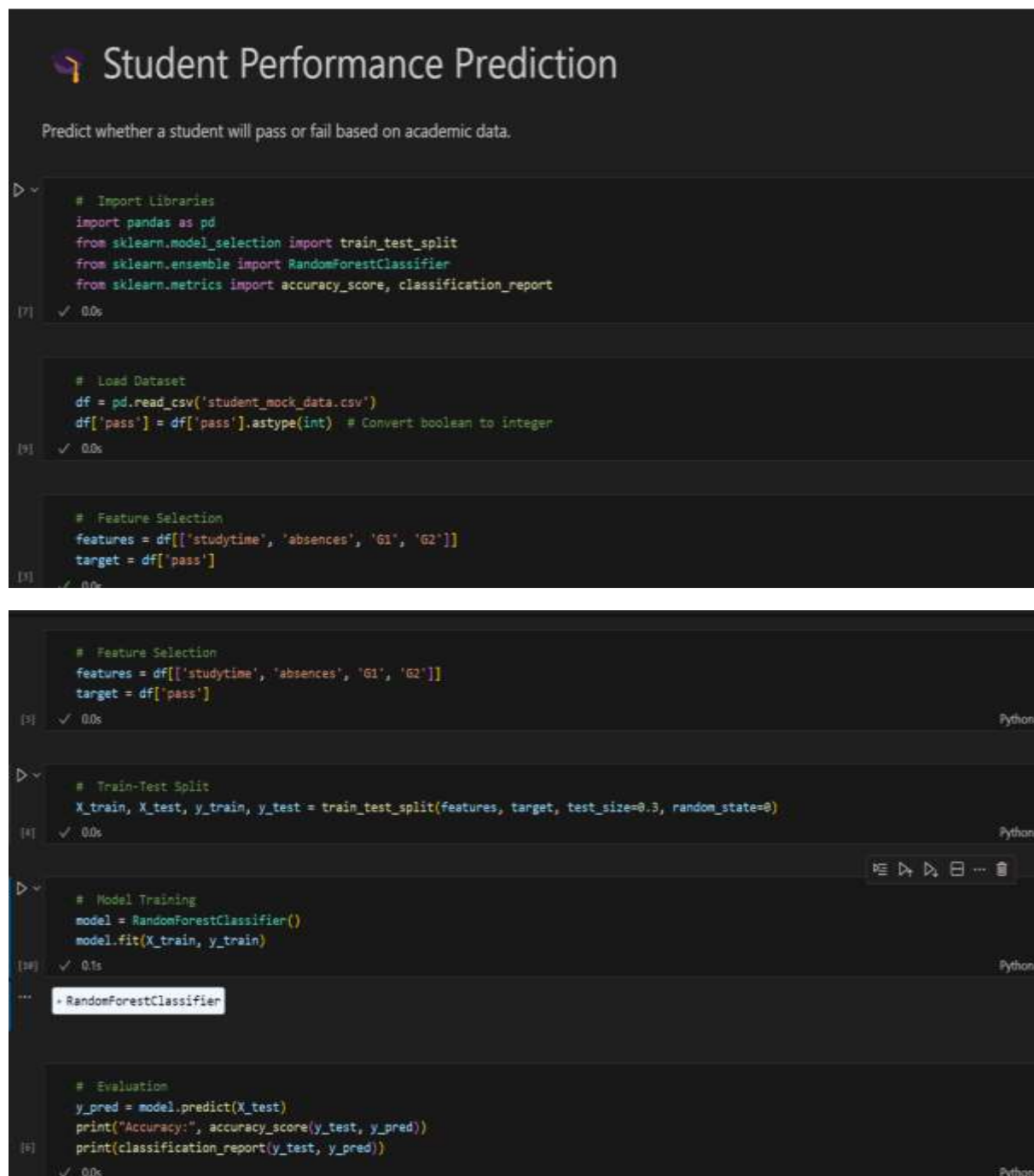


TASK -2

PREDICTIVE ANALYSIS USING MACHINE LEARNING

Data Source: Download student_mock_data.csv

Steps in the Jupyter Notebook:



The screenshot displays a Jupyter Notebook titled "Student Performance Prediction" with the subtitle "Predict whether a student will pass or fail based on academic data." The notebook contains several code cells for data analysis and machine learning.

```
# Import Libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Load Dataset
df = pd.read_csv('student_mock_data.csv')
df['pass'] = df['pass'].astype(int) # Convert boolean to Integer

# Feature Selection
features = df[['studytime', 'absences', 'G1', 'G2']]
target = df['pass']

# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.3, random_state=0)

# Model Training
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Evaluation
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

The notebook interface shows the following details:

- Cell 1:** Imports necessary libraries (pandas, sklearn). Execution time: 0.0s.
- Cell 2:** Loads the dataset and converts the 'pass' column to integer. Execution time: 0.0s.
- Cell 3:** Selects features and target variable. Execution time: 0.0s.
- Cell 4:** Splits the data into training and testing sets. Execution time: 0.0s.
- Cell 5:** Trains the RandomForestClassifier model. Execution time: 0.1s.
- Cell 6:** Evaluates the model by predicting on the test set and printing accuracy and classification report. Execution time: 0.0s.

The notebook also shows a variable inspector for the `RandomForestClassifier` object and a toolbar with icons for undo, redo, run, and other actions.

```
# Evaluation
y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

[6] ✓ 0.0s

... Accuracy: 0.5166666666666667

	precision	recall	f1-score	support
0	0.54	0.42	0.47	31
1	0.50	0.62	0.55	29
accuracy			0.52	60
macro avg	0.52	0.52	0.51	60
weighted avg	0.52	0.52	0.51	60

Student Performance Prediction – Highlights

- **Goal:** Predict if a student will pass or fail using their study habits and grades.
- **Machine Learning Type:** Classification using Random Forest Classifier.
- **Features Used:** Study time, absences, and first two grade scores (G1, G2).
- **Steps Included:**
 - Feature selection
 - Model training
 - Evaluation with accuracy & metrics
- **Outcome:** Helps identify at-risk students for early intervention