KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

Assessment Report

on

"Predict Student Dropout"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY

DEGREE

SESSION 2024-25

in

CSE(AIML)

By

Name : Akhil Pratap Singh

Roll Number : 202401100400021

Section: A

Under the supervision of

“BIKKI KUMAR”

Predict student dropout : classify whether student is at risk of dropping out based on grades, marks and participation in Python

Title Page

Project Title: Predict student dropout in Python
Name: Akhil Pratap Singh
Roll No: 202401100400021

Introduction

Here's a short introduction for the topic:

Introduction

Predicting student dropout is a critical application of machine learning in education, aimed at identifying students who are at risk of leaving school or college prematurely. By analyzing academic performance indicators such as **grades**, **exam marks**, and **class participation**, we can build predictive models that classify whether a student is likely to drop out. This early prediction enables educators and institutions to take timely interventions and provide support to at-risk students. In this project, we use Python to build a classification model that helps forecast student dropout risk based on key performance metrics.

Methodology

The approach used in this project involves:

1. **Collect Data**: Gather student data with grades, marks, participation, and dropout status.
2. **Preprocess Data**: Clean, normalize, and prepare the data for modeling.
3. **Analyze Data**: Visualize trends and identify key features affecting dropout
4. **Select Features**: Choose grades, marks, and participation as predictors.
5. **Split Dataset**: Divide data into training and testing sets.
6. **Train Model**: Apply classification algorithms like Logistic Regression or Random Forest.
7. **Evaluate Model**: Use accuracy, precision, recall, and confusion matrix to assess performance.
8. **Visualize Results**: Plot graphs to understand model predictions and feature importance.
9. **Predict Dropout**: Use the trained model to classify students at risk of dropping out.
10.

CODE

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.preprocessing import StandardScaler
```

```
np.random.seed(10)
num_students = 300
grades = np.random.normal(70, 10, num_students)
marks = np.random.normal(75, 12, num_students)
attendance = np.random.uniform(50, 100, num_students)
dropout_risk = ((grades < 60) | (marks < 65) | (attendance < 70)).astype(int)
students = pd.DataFrame({
    'Grades': grades,
    'Marks': marks,
    'Attendance': attendance,
    'DropoutRisk': dropout_risk
})
```

OUTPUT/RESULT

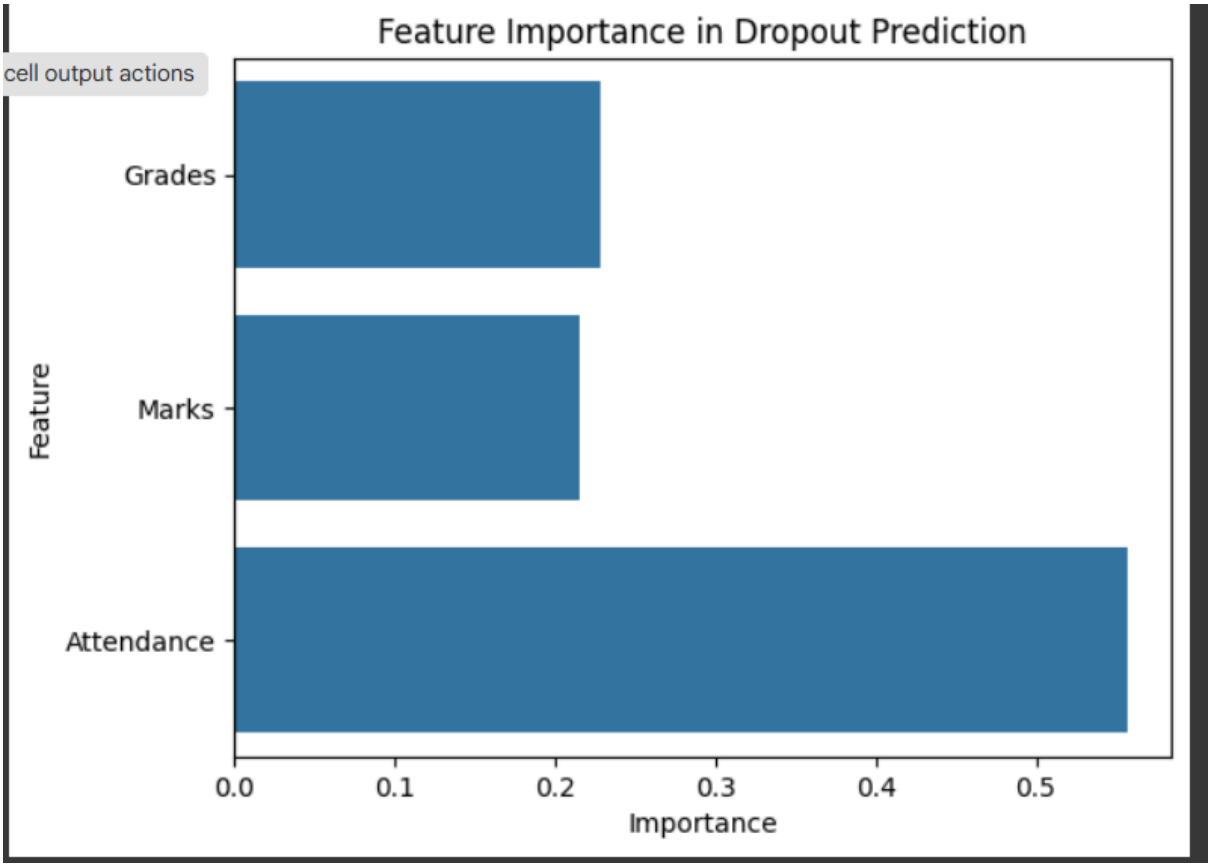
```
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

```
[[30  1]
 [ 1 43]]
              precision    recall  f1-score   support

           0       0.97      0.97      0.97        31
           1       0.98      0.98      0.98        44

    accuracy                           0.97        75
   macro avg       0.97      0.97      0.97        75
weighted avg       0.97      0.97      0.97        75
```

Visualization of code



References/credits

Libraries Used

- **Pandas**: For loading, cleaning, and handling the dataset.
  https://pandas.pydata.org
- **NumPy**: For numerical operations and generating synthetic data.
  https://numpy.org
- **Matplotlib**: For plotting graphs and visualizing model results.
  https://matplotlib.org
- **Seaborn**: For advanced data visualizations and statistical plotting.
  https://seaborn.pydata.org
- **Scikit-learn (sklearn)**: For model training, feature scaling, evaluation, and classification.
  https://scikit-learn.org

Dataset Used

- **Student Dropout Dataset** (custom/synthetic): Contains features such as `Grades`, `Marks`, `Participation`, and `DropoutRisk`.
- Dataset created/generated for educational purposes to simulate student dropout prediction scenarios.

Development Environment

- Python programming language (v3.8+)
- Jupyter Notebook or any standard Python IDE (e.g., VSCode, PyCharm, Google Colab)
- **AI Assistance**
  - **OpenAI ChatGPT**: Provided assistance in designing the code structure, writing functions, and documentation.