### Name: Avishkar Dinesh Chothwe

**Roll no:** **282029** **PRN :- 22310789**

**Batch:** B2

**Assignment 4**

**Statement:**

Q. Apply an appropriate ML algorithm on a dataset consisting of student admission details to predict the chance of admission. Create a confusion matrix based on the above data and find:

a) Accuracy

b) Precision

c) Recall

d) F1-score

**Objective:**

* To develop a machine learning model using **Logistic Regression** that predicts the binary outcome of a student's admission based on various academic and admission-related features.
* To evaluate the performance of the model using appropriate classification metrics like **accuracy**, **precision**, **recall**, **F1-score**, and **confusion matrix**.

**Resources used:**

* **Hardware**: Standard computer system capable of running Python.
* **Software & Libraries**:
  + Python
  + Jupyter Notebook
  + pandas
  + scikit-learn
  + matplotlib
  + seaborn
* **Dataset**: Admission\_Predict.csv, containing student admission data with attributes like GRE Score, TOEFL Score, CGPA, etc.

**Introduction to Machine Learning for Student Admission Prediction:**

In today’s data-driven world, machine learning (ML) has emerged as a powerful tool for making informed decisions based on patterns and trends in data. One of the practical applications of ML is in the field of college admission prediction, where historical admission data is analyzed to forecast the likelihood of an applicant being accepted into a program.

Admission decisions often rely on multiple factors such as GRE scores, TOEFL scores, undergraduate GPA, research experience, and letters of recommendation. Traditionally, evaluating these factors has been subjective and time-consuming. Machine learning offers a solution by using statistical algorithms that can learn from historical data and make accurate predictions for future applicants.

In this project, we apply Logistic Regression, a popular supervised machine learning algorithm, to predict whether a student will be admitted to a graduate program based on their profile. The model classifies students into two categories: likely to be admitted and not likely to be admitted, based on a threshold set for the "Chance of Admit" variable.

By automating this process, educational institutions can not only reduce manual workload but also enhance the fairness and transparency of admission decisions. For students, such predictive tools can provide insights into their chances of acceptance and help them make informed decisions about their academic future.

**Methodology:**

 **Data Loading**: Read the CSV file containing admission data using pandas.

 **Preprocessing**:

* Remove unnecessary columns like Serial No..
* Convert the regression output, Chance of Admit, into a binary classification problem.

 **Feature Selection**: Use relevant features such as GRE Score, TOEFL Score, CGPA, etc., as predictors.

 **Train-Test Split**: Divide the dataset into training and testing sets in an 80:20 ratio.

 **Model Building**: Use **Logistic Regression** from scikit-learn to fit the training data.

 **Prediction**: Predict admission outcomes on the test data.

 **Evaluation**:

* Generate a confusion matrix.
* Calculate accuracy, precision, recall, and F1-score.

 **Visualization**: Use seaborn and matplotlib to visualize the results.

**Confusion Matrix and Performance Metrics:**

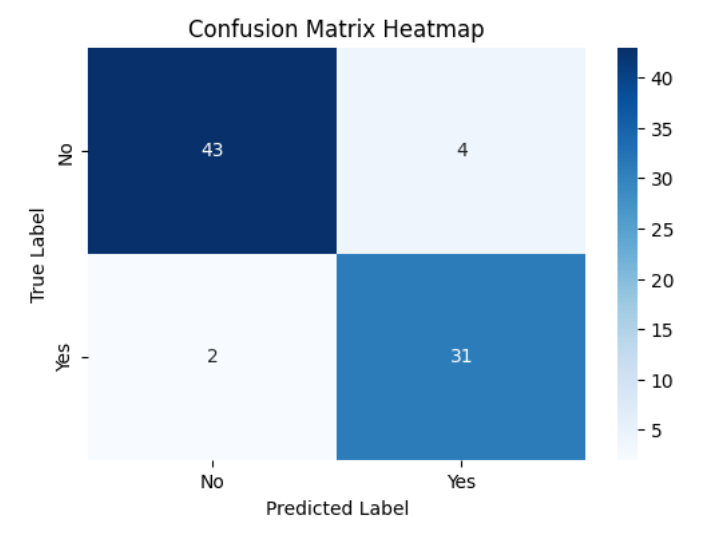
1. **Accuracy:** Measures the overall correctness of the model.
   * Formula: (TP + TN) / (TP + TN + FP + FN)
2. **Precision:** Indicates how many of the predicted positive cases were actually positive.
   * Formula: TP / (TP + FP)
3. **Recall (Sensitivity):** Measures how well the model identifies actual positive cases.
   * Formula: TP / (TP + FN)
4. **F1-score:** Harmonic mean of precision and recall, balancing both metrics.
   * Formula: 2 \* (Precision \* Recall) / (Precision + Recall)

**Advantages of ML in Admission Prediction:**

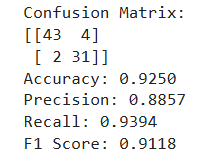
1. **Automation and Efficiency**  
   Machine learning automates the evaluation process, allowing institutions to process thousands of applications quickly and efficiently.
2. **Data-Driven Decisions**  
   ML models make decisions based on historical data and patterns, reducing bias and promoting consistency in the admission process.
3. **Scalability**  
   Once trained, the model can handle a large volume of applications without additional cost or time.
4. **Predictive Insights for Applicants**  
   Students can get an estimate of their admission chances, helping them set realistic expectations and improve weak areas.
5. **Customization and Continuous Learning**  
   Models can be fine-tuned for different institutions or programs and updated regularly as new data becomes available.

**Disadvantages:**

1. **Data Quality and Availability**  
   The accuracy of predictions heavily depends on the quality and completeness of the historical data used to train the model.
2. **Overfitting and Generalization**  
   A model may perform well on training data but poorly on new data if it overfits or doesn't generalize well.
3. **Lack of Transparency (Black Box Issue)**  
   Some ML algorithms (especially complex ones) may not provide clear reasoning behind predictions, making it hard to explain decisions.
4. **Ethical and Fairness Concerns**  
   If the training data contains biases (e.g., gender, ethnicity), the model might unintentionally learn and propagate those biases.
5. **Dependence on Quantitative Factors**  
   Machine learning models primarily rely on quantifiable features and may overlook qualitative factors like personal statements or recommendation letters.

**Results: **

|  |  |
| --- | --- |
|  |  |

****

**Conclusion:**

This project successfully demonstrates how machine learning, specifically logistic regression, can be used to predict college admissions based on student profiles. The model provides a fast, data-driven approach to decision-making, offering valuable insights for both institutions and applicants.

4o