Social Network Ads Classifier Project

Name: **Touseef Ahmed**

Objective

The aim of this project is to predict whether a user purchases a product based on their Age and Estimated Salary using three machine learning classification models:

- Naive Bayes

- K-Nearest Neighbors (KNN)

- Decision Tree

Dataset Information

Dataset name: Social_Network_Ads.csv

Source: https://www.kaggle.com/datasets/rakeshrau/social-network-ads

Features used: Gender (encoded as 0/1), Age, EstimatedSalary

Target variable: Purchased (0 = No, 1 = Yes)

Data Preprocessing

1. Dropped User ID column.

2. Encoded Gender using LabelEncoder.

3. Standardized Age and EstimatedSalary using StandardScaler.

4. Split data into 75% training and 25% testing sets.

Models Used

1. Gaussian Naive Bayes

2. K-Nearest Neighbors (k=3, 5, 7)

3. Decision Tree (Gini and Entropy criteria)

Evaluation Metrics

Accuracy, Precision, Recall, F1 Score, and Confusion Matrix were used.

Model Performance Summary

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.89	0.89	0.89	0.89
KNN (k=3)	0.90	0.90	0.90	0.90
KNN (k=5)	0.91	0.91	0.91	0.91
KNN (k=7)	0.89	0.89	0.89	0.89
Decision Tree (Gini)	0.89	0.89	0.89	0.89
Decision Tree (Entropy)	0.90	0.90	0.90	0.90

Conclusion

Best Performing Model: K-Nearest Neighbors with k=5.
It achieved the highest performance across all metrics.
Balanced accuracy and generalization without overfitting.
Decision Trees performed comparably but risk overfitting.
Naive Bayes was simplest but slightly less effective.