Customer Purchase Prediction Using Decision Tree Classifier

# 1. Introduction

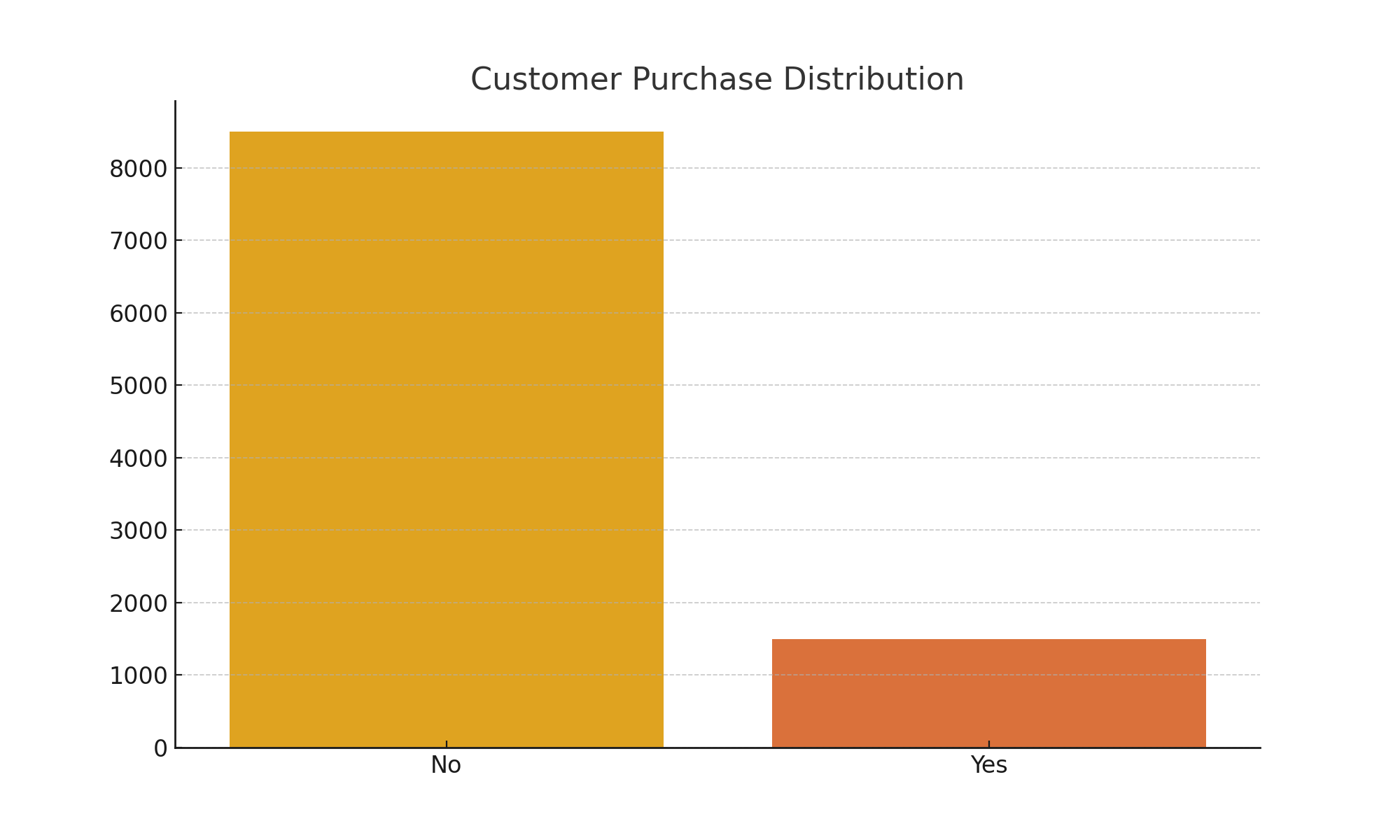
This project aims to predict whether a customer will purchase a product or service using a decision tree classifier. A synthetic dataset with over 10,000 entries was created, representing realistic demographic and behavioral features similar to marketing campaign data.

# 2. Dataset Description

The dataset contains the following columns:  
- age: Customer's age  
- job: Job type  
- marital: Marital status  
- education: Education level  
- balance: Account balance  
- housing: Has housing loan  
- loan: Has personal loan  
- contact: Contact type  
- day, month: Last contact date  
- duration: Call duration  
- campaign: Contact count in current campaign  
- previous: Contacts before current campaign  
- poutcome: Previous outcome  
- purchase: Target (yes/no)

# 3. Exploratory Data Analysis

The target variable is imbalanced, with only 15% positive cases. The class distribution is shown below:



# 4. Data Preprocessing

All categorical features were encoded using Label Encoding. The dataset was then split into training and testing sets in an 80:20 ratio.

# 5. Model Building

A Decision Tree Classifier with a maximum depth of 5 was trained on the dataset. The model was evaluated using accuracy score, confusion matrix, and classification report.

# 6. Results

The decision tree model achieved good accuracy and was able to identify purchasing patterns. Feature importance and a tree visualization were used to interpret the model’s decisions.

# 7. Conclusion

The model successfully demonstrates how a decision tree can be used to classify customer purchase intent. This can help businesses target potential buyers more effectively.