

Random Forest

1. Objective

Build a Random Forest model to predict whether a customer will buy a product based on age and salary.

2. Install Required Libraries (One Time Only)

```
pip install numpy pandas matplotlib scikit-learn
```

3. Import Libraries

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

4. Create Sample Dataset

```
data = pd.DataFrame({
    "Age": [22,25,47,52,46,56,55,60,62,23],
    "Salary": [20000,25000,50000,60000,52000,58000,57000,65000,70000,22000],
    "Buy": [0,0,1,1,1,1,1,1,1,0]
})

print(data)
```

5. Separate Features and Target

```
X = data[["Age","Salary"]]
y = data["Buy"]
```

6. Split Train and Test Data

```
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)
```

7. Create Model

```
model = RandomForestClassifier(n_estimators=100)
```

8. Train Model

```
model.fit(X_train, y_train)  
print("Model Trained Successfully")
```

9. Make Predictions

```
y_pred = model.predict(X_test)  
print("Actual:", list(y_test))  
print("Predicted:", list(y_pred))
```

10. Evaluate Model

```
accuracy = accuracy_score(y_test, y_pred)  
print("Accuracy:", accuracy)
```

```
print("Confusion Matrix:")  
print(confusion_matrix(y_test, y_pred))
```

```
print("Classification Report:")  
print(classification_report(y_test, y_pred))
```

11. Predict New Customer

```
new_customer = [[40, 48000]]
prediction = model.predict(new_customer)

if prediction[0] == 1:
    print("Customer WILL BUY")
else:
    print("Customer WILL NOT BUY")
```

12. Key Learning

- Random Forest is ensemble of multiple Decision Trees
 - Reduces overfitting
 - More accurate than single tree
- Works for classification and regression

Random Forest is an ensemble supervised learning algorithm that builds multiple decision trees and combines their outputs for better accuracy.