EXPERIMENT 10

CO5: To implement and evaluate predictive models, including classification and machine learning models.

Aim: To conduct a Decision Tree analysis on a sample dataset and understand the process of classification using features and target values.

Objective
- To learn how to construct and interpret a Decision Tree.
- To classify data based on various features and analyze the influence of each feature on the decision.
- To visualize Decision Tree results in Origin Pro 2024 for further analysis.

Theory
A Decision Tree is a machine learning tool that allows for classification and regression. It is a tree-like model where nodes represent decisions based on certain conditions, and branches represent possible outcomes.

Key Concepts in Decision Trees:
1. **Root Node**: Represents the entire dataset and splits based on the feature providing the best classification.
2. **Decision Nodes**: Intermediate nodes representing decisions on feature values.
3. **Leaf Nodes**: Terminal nodes representing classification outcomes.
4. **Splitting Criteria**: Criteria like **Gini Index** and **Entropy** are used to decide the best split.
5. **Pruning**: The process of removing less informative branches to avoid overfitting.

Advantages of Decision Trees:
- Simple to understand and interpret.
- Useful for both numerical and categorical data.

Disadvantages:
- Prone to overfitting, especially with noisy data.
- Can be sensitive to changes in the data.

Procedure

Step 1: Prepare the Dataset
- Start by creating a sample dataset. This dataset will include 10 rows with various features like **Age**, **Income**, **Credit Rating**, and the target variable **Loan Approved**.

Age	Income	Credit Rating	Loan Approved
25	High	Good	Yes
30	Low	Bad	No
45	Medium	Good	Yes
40	Low	Good	No
22	Medium	Bad	No
35	High	Good	Yes
28	Low	Good	Yes
50	Medium	Bad	No
32	High	Good	Yes
55	Low	Bad	No

Step 2: Export the Dataset from Origin Pro
- Enter this data into an Origin worksheet.
- Export the data as a .csv file to use it in Python. In Origin, go to **File > Export > ASCII Options** and save the file.

Step 3: Run Decision Tree Analysis in Python
- Open a Python environment (Anaconda, Jupyter Notebook, or Google Colab) with scikit-learn installed.
- Use the following Python code to analyze the dataset and create a Decision Tree model:

python
Copy code

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv("path_to_file.csv")

# Prepare features and target variable
X = pd.get_dummies(data[['Age', 'Income', 'Credit Rating']])
y = data['Loan Approved'].map({'Yes': 1, 'No': 0})

# Initialize and fit the Decision Tree model
model = DecisionTreeClassifier(criterion="entropy", max_depth=3)
model.fit(X, y)

# Plot the Decision Tree
plt.figure(figsize=(12, 8))
plot_tree(model, feature_names=X.columns, class_names=['No', 'Yes'], filled=True)
plt.savefig("decision_tree_output.png")
plt.show()
```

- This code trains a Decision Tree using the sample data, and the plot_tree function will create a visual representation of the tree.
- Save the output as decision_tree_output.png.

Step 4: Import Results into Origin Pro
- Use **File > Import** in Origin Pro to load the decision_tree_output.png image.
- You can include this image in your report within Origin for documentation.

Step 5: Additional Analysis and Interpretation in Origin Pro
- After importing the Decision Tree plot, you can annotate or analyze it further in Origin Pro.
- Use Origin's text tools to add observations or highlight key decision points on the imported image.

Results: The Decision Tree visualization will show the classification based on Age, Income, and Credit Rating. Each path from the root node to a leaf node represents a series of decisions leading to the final classification (e.g., loan approval or rejection). By analyzing the tree, you can determine which features have the most significant impact on the classification.

Example Observations:
- Applicants with high income and a good credit rating are likely to have their loan approved.
- Low-income applicants with a bad credit rating are generally denied loans.

Learning Outcomes:
- You have learned how to set up a Decision Tree analysis, including splitting criteria and interpreting outcomes.

- You became familiar with using Python to conduct machine learning analysis and Origin Pro for data visualization.
- You developed an understanding of the impact of each feature on the classification, enabling more informed decision-making.

Screenshots: From Jupyter Notebook

```python
import pandas as pd
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv("C:/Users/kaush/Downloads/Sheet1.csv")

# Prepare features and target variable
X = pd.get_dummies(data[['Age', 'Income', 'Credit Rating']])
y = data['Loan Approved'].map({'Yes': 1, 'No': 0})

# Initialize and fit the Decision Tree model
model = DecisionTreeClassifier(criterion="entropy", max_depth=3)
model.fit(X, y)

# Plot the Decision Tree
plt.figure(figsize=(12, 8))
plot_tree(model, feature_names=X.columns, class_names=['No', 'Yes'], filled=True)
plt.savefig("decision_tree_output.png")
plt.show()
```
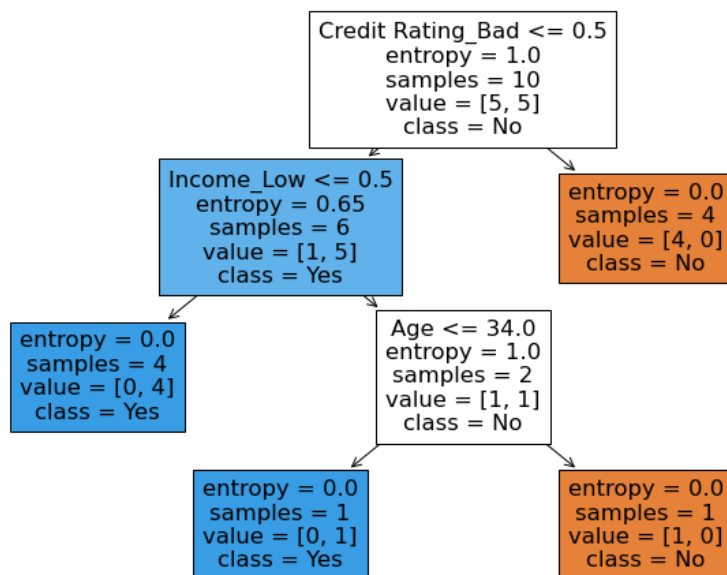
From Origin Pro Software

Book1 - Experiment-10_Data.xlsx

	A(X)	B(Y)	C(Y)	D(Y)
Long Name	Age	Income	Credit Rating	Loan Approved
Units				
Comments				
F(x)=				
Sparklines				
1	25	High	Good	Yes
2	30	Low	Bad	No
3	45	Medium	Good	Yes
4	40	Low	Good	No
5	22	Medium	Bad	No
6	35	High	Good	Yes
7	28	Low	Good	Yes
8	50	Medium	Bad	No
9	32	High	Good	Yes
10	55	Low	Bad	No
11				
12				
13				

Sheet1

Image1 - decision_tree_output.png

864x576 pixels; 1/1; (858,511), RGBA(255,255,255,0)