A Project Report

On

Decision Tree

conducted as part of the course CSE CS3131 Submitted by

Pradyumna Singh 209301309

Ansh Ujlayan 209301050

Akansha Ghosh 209301524

in partial fulfilment for the award of the degreeof

BACHELOR OF TECHNOLOGY

In

Computer Science & Engineering

Department of Computer Science & Engineering,

School of Computing and IT,

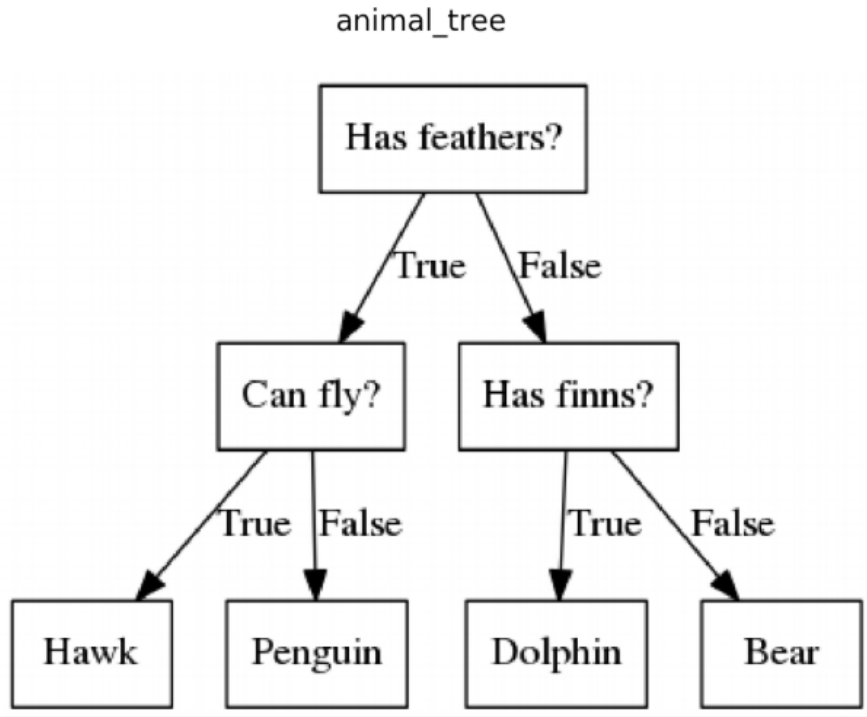
Manipal University Jaipur,

November 2022

PROBLEM STATEMENT

A decision tree is one of the supervised machine learning algorithms. This algorithm can be used for regression and classification problems-yet, is mostly used for classification problems. A decision tree follows a set of if-else conditions to visualize the data and classify it according to the conditions.

Uncertain values can lead to complex calculations and uncertain outcomes; decision trees are unstable, and minor data changes can lead to major structure changes; information gain in decision trees can be biased; and decision trees can often be relatively inaccurate.



AIM AND OBJECTIVE

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data). In Decision Trees, for predicting a class label for a record we start from the root of the tree.

Let's say we have a problem to predict whether a customer will pay his renewal premium with an insurance company (yes/ no). Here we know that the income of customers is a significant variable but the insurance company does not have income details for all customers. Now, as we know this is an important variable, then we can build a decision tree to predict customer income based on occupation, product, and various other variables. In this case, we are predicting values for the continuous variables.

MOTIVATION

- Expansion of knowledge in the domain of AI and Machine Learning.
- Improve one's capability for an overall development incompetitive and algorithmic skills of an individual.
- Using the attained knowledge as a medium that enables aperson to be self-reliant in the domain of AI.
- Equipping ourselves in these ways that's helps a persons inways more than one.

Algorithm

1. Find the best attribute and place it on the root node of the tree.

2. Now, split the training set of the dataset into subsets. While making the subset make sure that each subset of training dataset should have the same value for an attribute.

3. Find leaf nodes in all branches by repeating 1 and 2 on each subset.

Libraries Used

1. Numpy
2. Matplotlib.pyplot
3. Pandas
4. Math
5. Copy

Working example Illustration

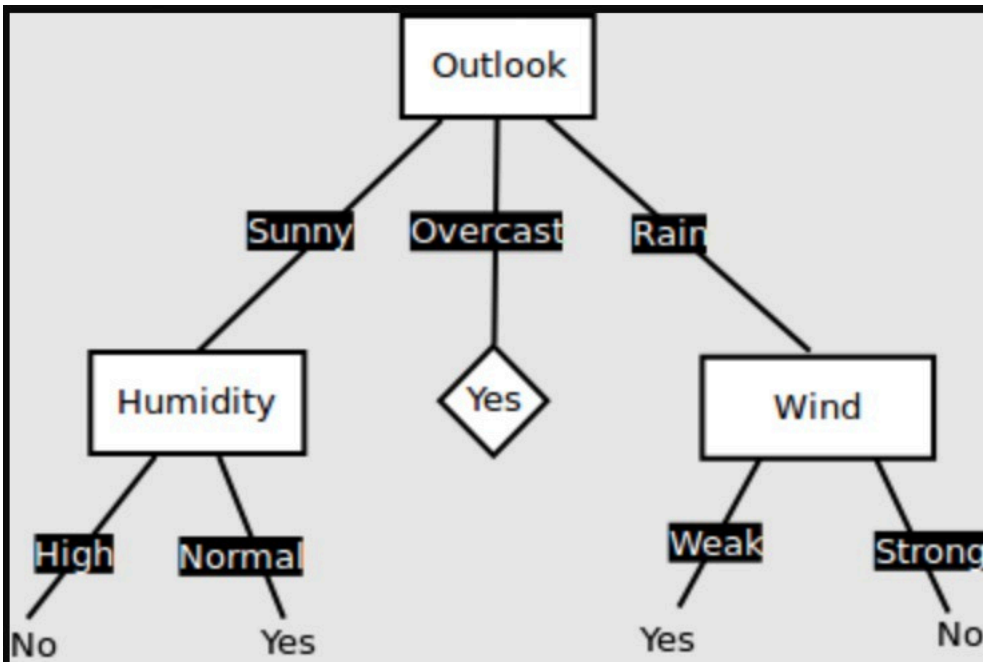
Input file:

---

tennis

day	outlook	temp	humidity	wind	play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Output:



main.py tennis.csv

```
            temp.decision = key
            root.childs.append(temp)
        return root


def traverse(root):
    print(root.decision)
    print(root.value)

    n = len(root.childs)
    if n > 0:
        for i in range(0, n):
            traverse(root.childs[i])
```

input

```
Start
outlook
Sunny
humidity
High
No
Normal
Yes
Overcast
Yes
Rain
wind
Weak
Yes
Strong
No


...Program finished with exit code 0
Press ENTER to exit console.
```