MENTORNESS ARTICLE
Task 1

Batch Name: MIP-ML-08

Machine Learning Algorithms Overview: A Comprehensive Review of Decision Trees

By Md. Harun-Or-Rashid Khan

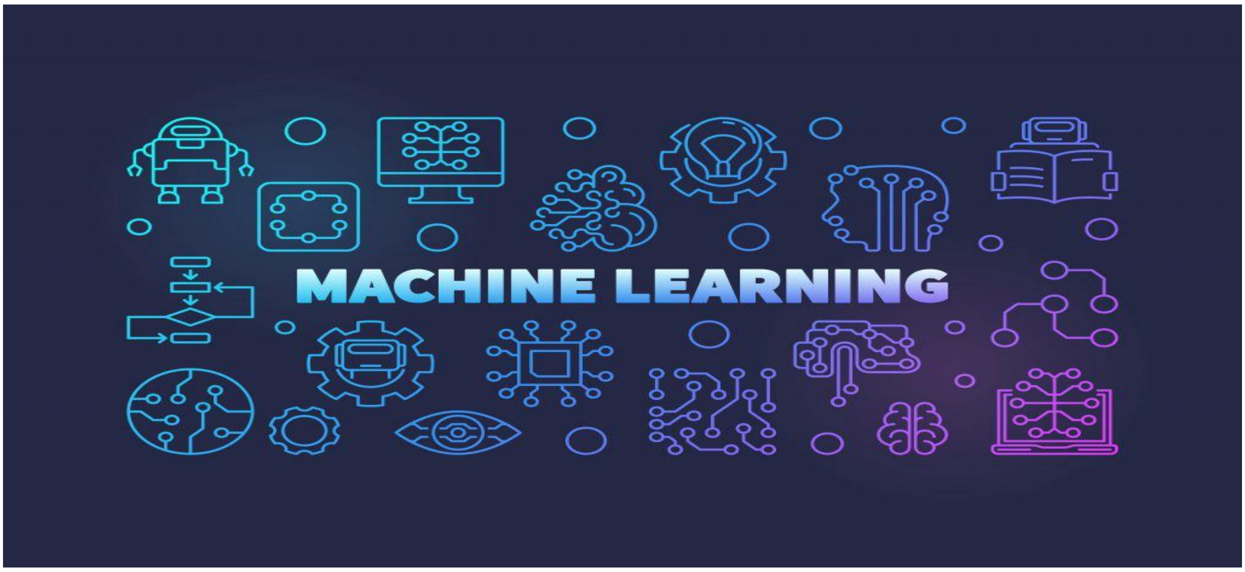


Fig. 1. Conceptual figure of machine learning.

Introduction:

In the dynamic landscape of artificial intelligence and machine learning, decision trees have become the backbone of innovation, providing intuitive solutions to complex decision-making problems. As we explore the ever-expanding field of intelligent systems, understanding the complexity of decision trees becomes critical. These versatile algorithms not only provide actionable insights but also pave the way for interpretable and transparent models. In this comprehensive review, we embark on a journey to uncover the depth of decision trees, exploring their fundamentals, applications across domains, and recent advances. Through a combination of theory, practical examples, and real-world use cases, we aim to provide both beginners and experienced practitioners with the necessary knowledge and insights to harness the transformative potential of decision trees in the pursuit of excellence in artificial intelligence.

Decision Trees:

Imagine you're faced with the decision of whether to go out and play based solely on the weather. Decision trees work much like how you navigate the decision-making process. Initially, you will assess current weather conditions and make a decision based on a series of questions and considerations. If it's raining, you might choose to stay indoors; if it's sunny, you might go outside. However, if the sky is cloudy, you may further evaluate factors such as temperature and wind before making a final decision. Each step in this iterative process affects subsequent operations, reflecting the essence of the decision tree algorithm.

Weather	Temperature	Wind	Play
Sunny	Hot	Weak	No
Sunny	Hot	Strong	No
Cloudy	Hot	Weak	Yes
Rain	Cool	Weak	Yes
Cloudy	Cool	Strong	Yes
Cloudy	Mild	Strong	Yes
Rain	Cool	Strong	No
Rain	Mild	Weak	Yes

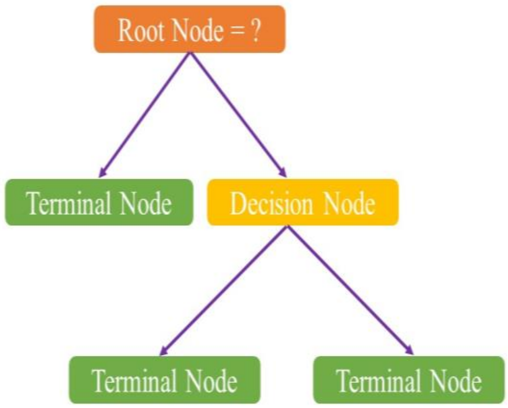


Fig. 2. Data Set and flowchart of a decision trees machine learning model.

You start by asking a fundamental question: Which feature serves as the root node in the decision tree? Using Python, we can easily find this critical piece of information with the code provided below:

```
# Convert categorical variables into numerical values using one-hot
encoding
df_encoded = pd.get_dummies(df[['Weather', 'Temperature','Wind']])

# Define target variable
y = df['Play']

# Create and train the decision tree classifier
clf = DecisionTreeClassifier()
```

```
clf.fit(df_encoded, y)

# Extract the feature at the root node
root_node_index = clf.tree_.feature[0]
root_node_feature = df_encoded.columns[root_node_index]

# Extract the actual feature name without the binary value
root_node_feature = root_node_feature.split('_')[0]

print("The first feature to start in the decision tree is:", root_node_feature)
```

```
The first feature to start in the decision tree is: Weather
```

Now we can delve into its mathematical intuition to understand how decision nodes are determined, employing concepts such as Entropy, Information Gain, and Gini Impurity. These metrics play pivotal roles in decision tree algorithms, guiding the process of feature selection and node splitting. Let's explore each of these concepts to gain a deeper understanding of the decision-making process within decision trees.

- **Entropy:** Entropy is a measure of randomness or uncertainty in a dataset. In the context of decision trees, it quantifies the impurity of a collection of examples. A dataset with low entropy means it is predominantly composed of examples from a single class, while a dataset with high entropy contains examples from multiple classes in similar proportions.

- **Information Gain:** Information Gain is a metric used to determine the effectiveness of splitting a dataset based on a particular feature. It measures the reduction in entropy (or increase in purity) that results from splitting the dataset using that feature. Higher Information Gain indicates that splitting the dataset based on that feature leads to more homogenous subsets in terms of class labels.

- **Gini Impurity:** Gini Impurity is another measure of impurity similar to entropy but calculated differently. It quantifies the probability of incorrectly classifying a randomly chosen element if it were randomly labeled according to the distribution of labels in the dataset. In decision trees, Gini Impurity is used as a criterion for determining the quality of a split, with lower Gini Impurity indicating a more homogeneous subset of data.

Mathematical Calculation:

First Calculate the information gain of 'Weather' feature:
We consider 'Yes' as '+' and 'No' as '-' and the target column probability of play is $\frac{4}{7}$ and not play is $\frac{3}{7}$.

Entropy of entire dataset: $(+5,-3) = -\frac{5}{8}\log_2\frac{5}{8} - \frac{3}{8}\log_2\frac{3}{8} = 0.95443$

Now calculate of all attributes:

Entropy of $Sunny_2(+0,-2) = 0$
In 'Weather' futures Sunny has 2 times w.r.to Play it's 'No'.
Similarly,
Entropy of $Cloudy_3(+3,-0) = 0$
Entropy of $Rain_3(+2,-1) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918296$

Information Gain of 'Weather'
$= \textbf{Entropy of entire dataset} - \frac{2}{8} \times \text{Entropy of } Sunny - \frac{3}{8} \times$ Entropy of $Cloudy - \frac{3}{8} \times$ Entropy of $Rain$
$= 0.61$

Secondly, Calculate the information gain of 'Temperature' feature:
Entropy of $Hot_3(+1,-2) = 0.9183$
Entropy of $Cool_3(+2,-1) = 0.9183$
Entropy of $Mild_2(+2,-0) = 0$

Information Gain of 'Temperature'
$= \textbf{Entropy of entire dataset} - \frac{3}{8} \times \text{Entropy of } Hot - \frac{3}{8} \times \text{Entropy of } Cool -$ $\frac{2}{8} \times$ Entropy of $Mild$
$= 0.2657$

Thirdly, Calculate the information gain of 'Wind' feature:
Entropy of $Weak_4(+3,-1) = 0.81128$
Entropy of $Cool_4(+2,-2) = 1$

Information Gain of ‘Wind’

$= \textbf{Entropy of entire dataset} - \frac{4}{8} \times \text{Entropy of } Weak - \frac{4}{8} \times \text{Entropy of } Cool$

$= 0.0488$

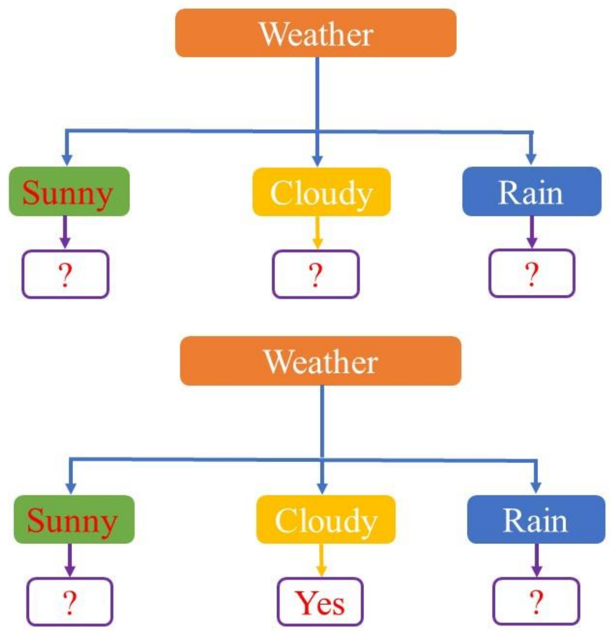
Summary of the three features are:

Information Gain of ‘Weather’ = 0.61

Information Gain of ‘Temperature’ = 0.2657

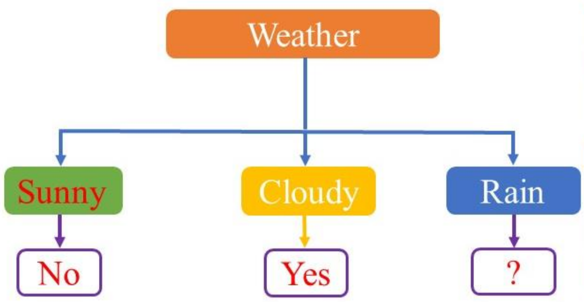
Information Gain of ‘Wind’ = 0.0488

Above three “Information Gain of ‘Weather’ is maximum. So, Decision or root node is ‘Weather’.



Weather	Temperature	Wind	Play
Sunny	Hot	Weak	No
Sunny	Hot	Strong	No
Cloudy	Hot	Weak	Yes
Rain	Cool	Weak	Yes
Cloudy	Cool	Strong	Yes
Cloudy	Mild	Strong	Yes
Rain	Cool	Strong	No
Rain	Mild	Weak	Yes

Fig. 3. Flowchart of decision trees with identify its decision node.



Weather	Temperature	Wind	Play
Sunny	Hot	Weak	No
Sunny	Hot	Strong	No

Weather	Temperature	Wind	Play
Rain	Cool	Weak	Yes
Rain	Cool	Strong	No
Rain	Mild	Weak	Yes

Fig. 4. Flowchart and dataset of decision trees with identify its decision and root nodes.

Now Calculate the information gain of ‘Rain’ of ‘Weather’ feature:

Entropy of Rain: $(+2, -1) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.91829$

Now calculate of all attributes:

Entropy of $Cool_2(+1, -1) = 1$

Entropy of $Mild_1(+1, -0) = 0$

Information Gain of ‘Temperature’

$= \textbf{Entropy of Rain} - \frac{2}{3} \times \text{Entropy of } Cool - \frac{1}{3} \times \text{Entropy of } Mild$

$= 0.25162$

Secondly, Calculate the information gain of ‘Wind’ feature:

Entropy of $Weak_2(+2, -0) = 0$

Entropy of $Strong_1(+0, -1) = 0$

Information Gain of ‘Wind’

$= \textbf{Entropy of Rain} - \frac{2}{3} \times \text{Entropy of } Weak - \frac{1}{3} \times \text{Entropy of } Strong$

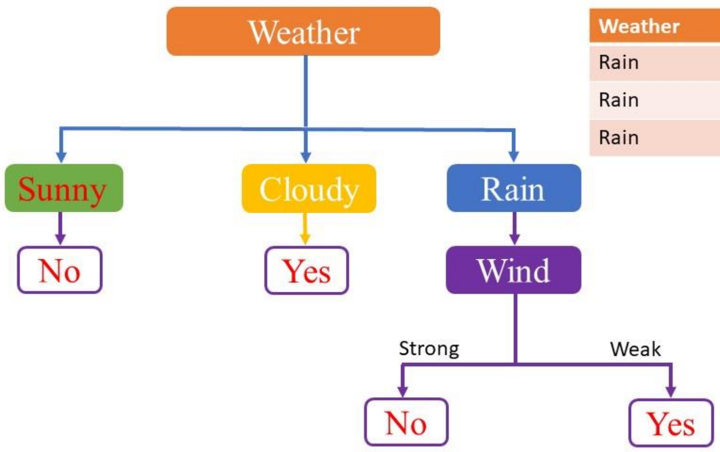
$= 0.91823$

Summary of the two features with respect to ‘Rain’:

Information Gain of ‘Temperature’ = 0.25162

Information Gain of ‘Wind’ = 0.91823

Above two “Information Gain of ‘Wind’ is maximum. So, next node is ‘Wind’.



Weather	Temperature	Wind	Play
Rain	Cool	Weak	Yes
Rain	Cool	Strong	No
Rain	Mild	Weak	Yes

Wind	Play
Weak	Yes
Strong	No
Weak	Yes

fig.5. Flowchart of a decision trees.

Conclusion:

In summary, decision trees are powerful tools in the machine-learning arsenal, offering a unique combination of simplicity, interpretability, and effectiveness. From basic tree-based models to advanced ensemble techniques, they continue to define the future of intelligent systems. As we move further into the world of artificial intelligence, let's recognize the critical role decision trees play in driving innovation, promoting understanding, and enabling informed decision-making. Through continued research, collaboration, and experimentation, decision trees are expected to remain at the forefront of machine learning and shape the future of future intelligent systems.