



DECISION TREES IN MACHINE LEARNING

By Rishabh Dubey





Quiz Time

QUICK REVISION QUIZ (MCQS & TRUE/FALSE)

Question 1: What does Linear Regression primarily do?

- A) Classifies data into categories
- B) Finds relationships between variables
- C) Clusters similar data points
- D) Encrypts data

QUICK REVISION QUIZ (MCQS & TRUE/FALSE)

Question 2: What is the equation for simple linear regression?

- ☐ A) $y = mx + b$
- ☐ B) $y = ax^2 + bx + c$
- ☐ C) $y = m_1x_1 + m_2x_2 + c$
- ☐ D) $y = \frac{x}{m} + b$

QUICK REVISION QUIZ (MCQS & TRUE/FALSE)

Question 3: How does Multiple Linear Regression differ from Simple Linear Regression?

- A) Uses more than one independent variable
- B) Uses non-linear relationships
- C) Only works for categorical data
- D) Requires deep learning

QUICK REVISION QUIZ (MCQS & TRUE/FALSE)

Question 4: Which metric is commonly used to evaluate a linear regression model?

- A) Accuracy
- B) Mean Squared Error (MSE)
- C) Recall
- D) Entropy



Introduction to Decision Trees

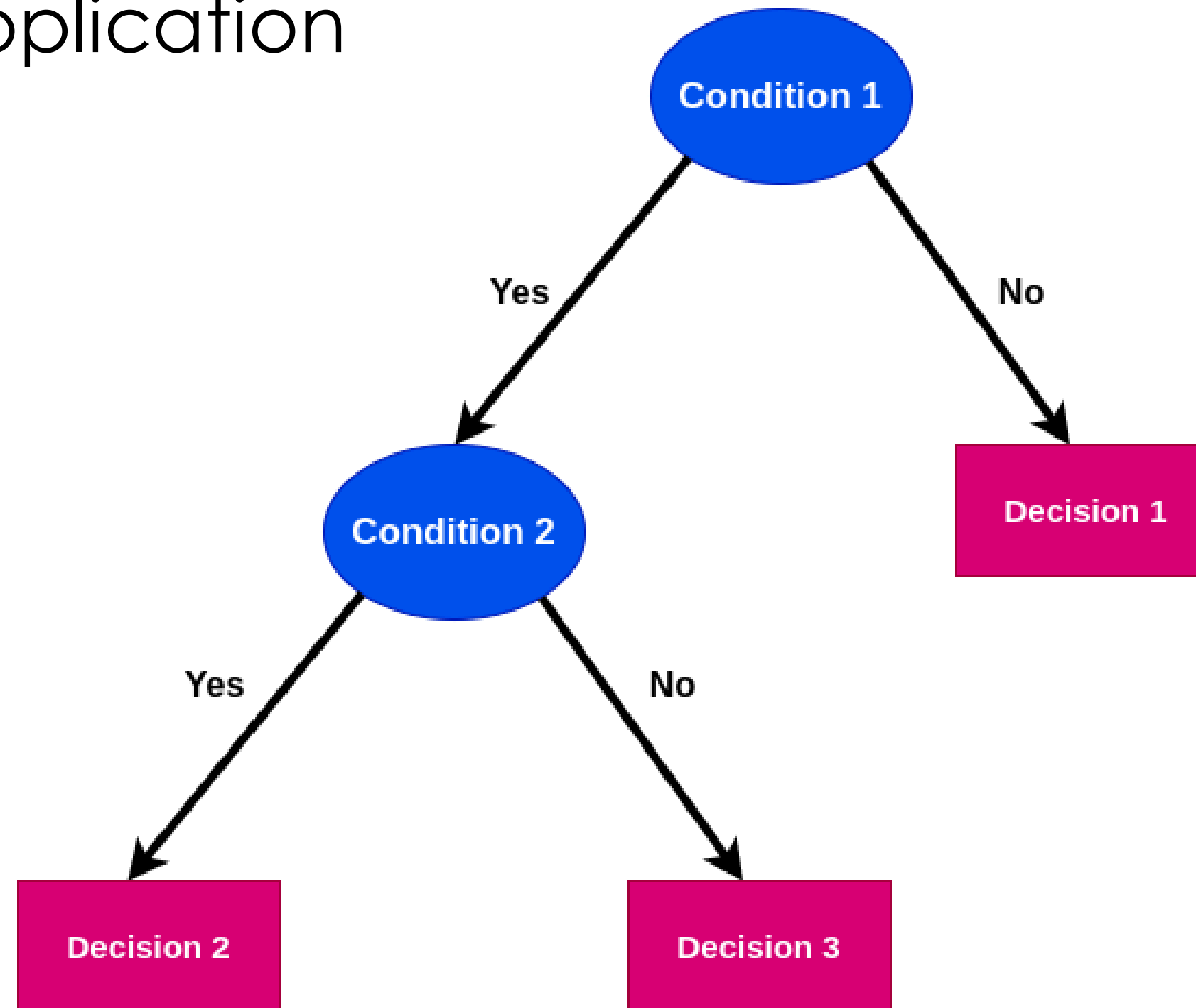
WHAT IS A DECISION TREE?

What is a Decision Tree?:

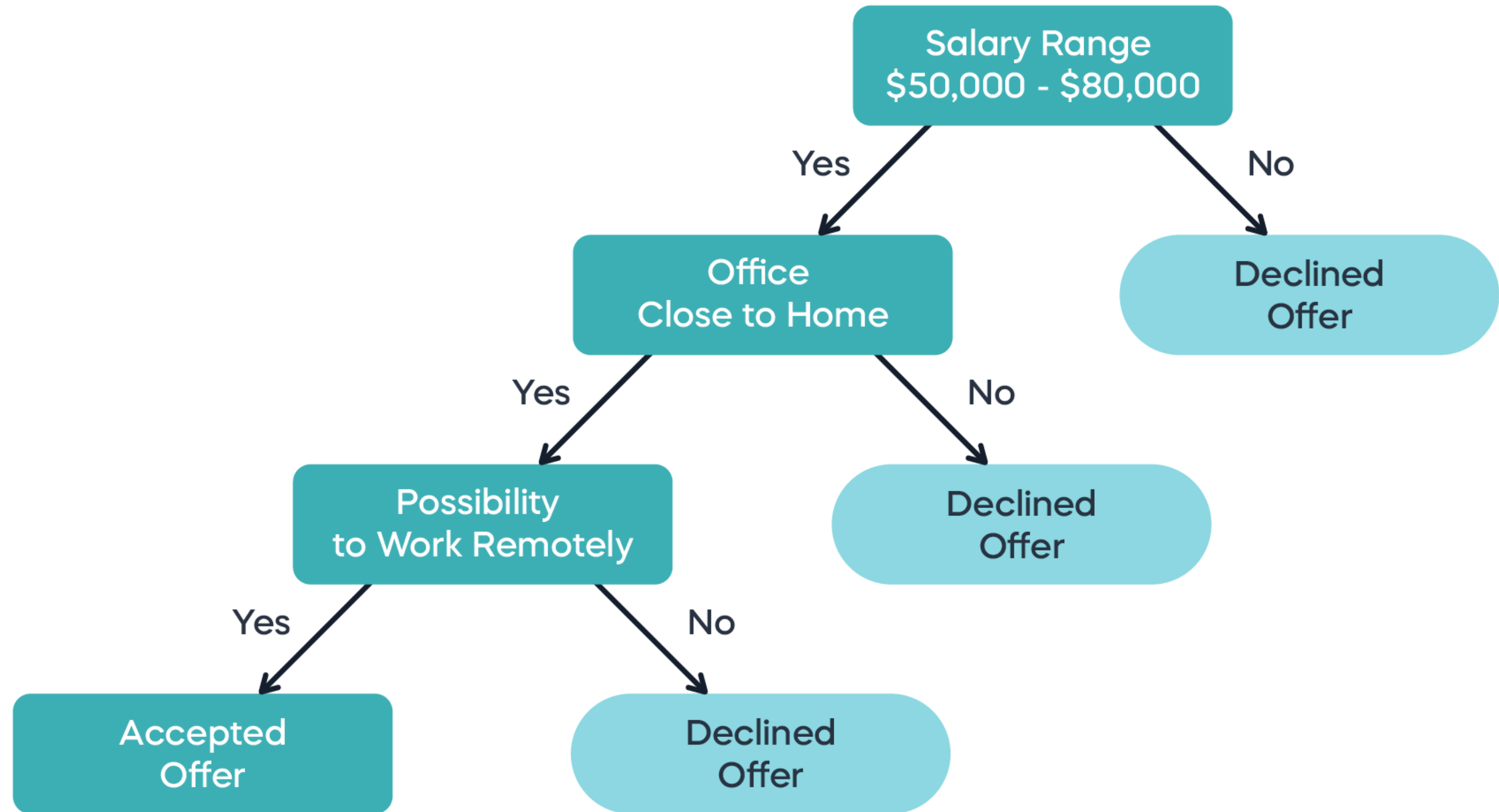
- A Decision Tree is a flowchart-like model that helps make predictions based on a sequence of decisions.
- Used for classification (Yes/No decisions) and regression (continuous values).
- Example: Should I play outside today? (Depends on weather conditions).

DECISION TREES IN MACHINE LEARNING





Understanding how decision trees make predictions and their real-world application



DECISION TREES IN MACHINE LEARNING

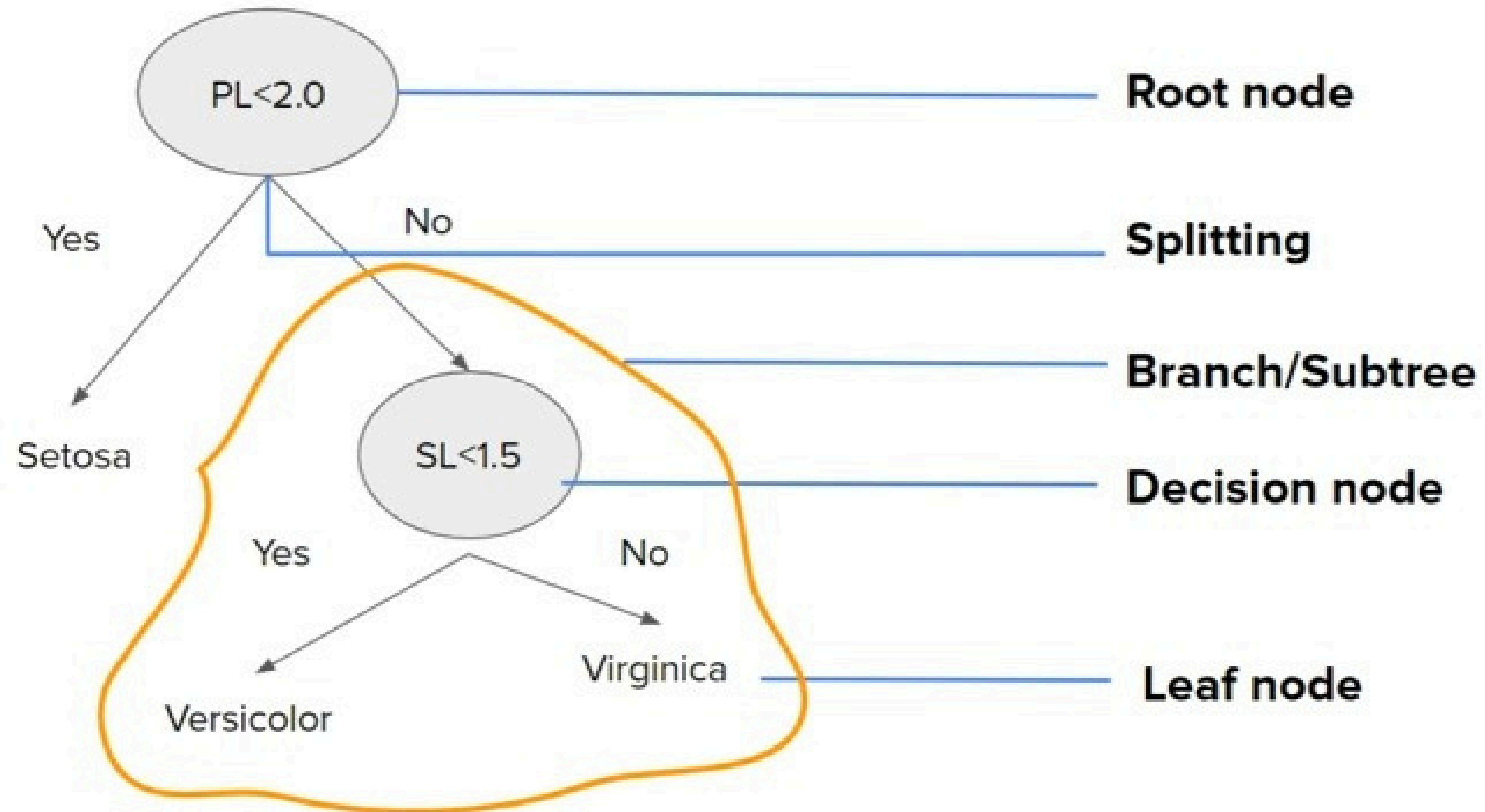


REAL-WORLD APPLICATIONS OF DECISION TREES

- Medical Diagnosis  – Predict diseases based on symptoms
- Loan Approval  – Bank checks credit history, income, etc.
- Customer Segmentation  – Classifying users based on behavior
- Spam Filtering  – Identifying spam emails

STRUCTURE OF A DECISION TREE

Terminology



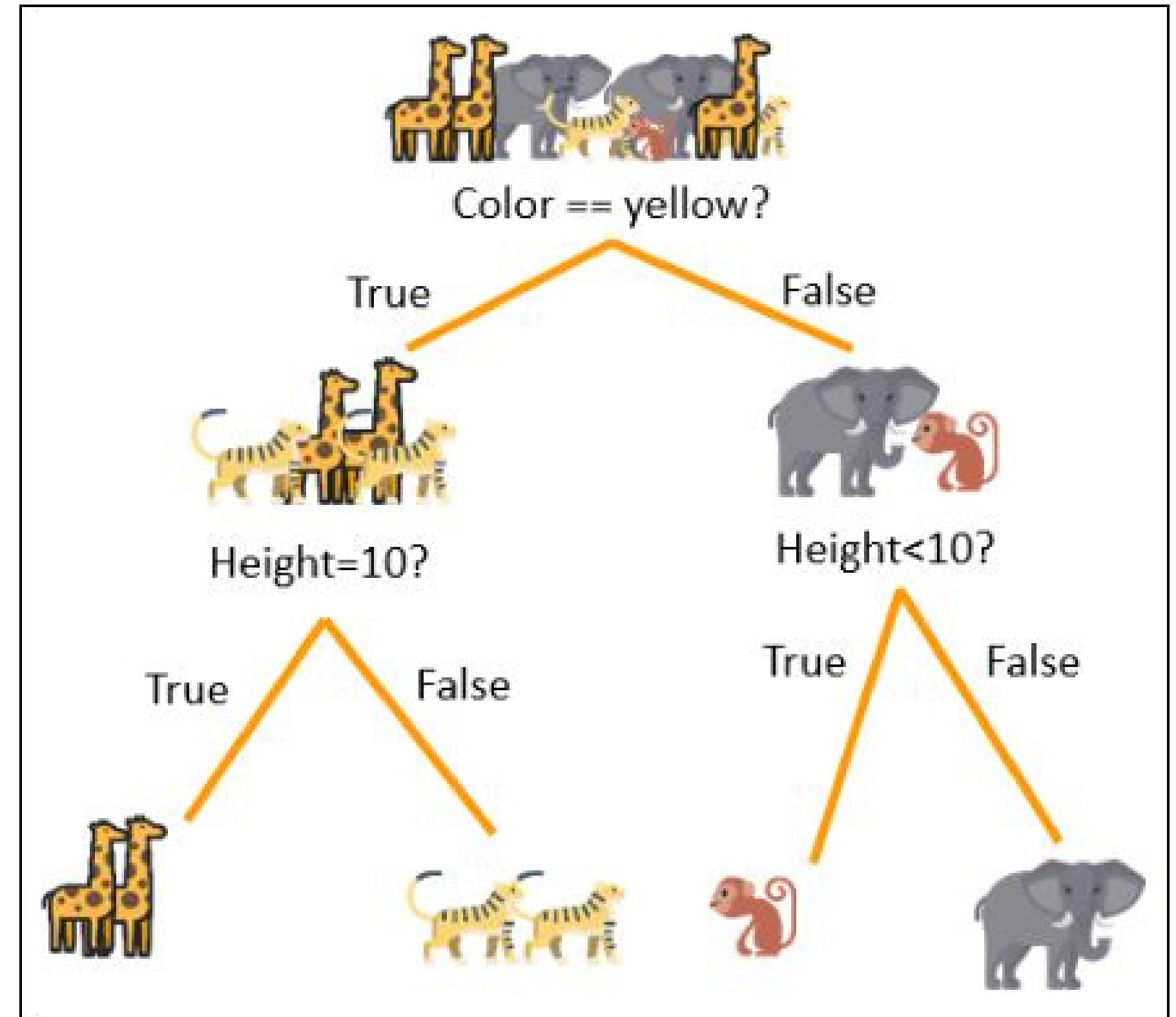


How Decision Trees Work?



HOW A DECISION TREE MAKES DECISIONS

- Each node represents a question based on feature values.
- The dataset is split into smaller groups based on these questions.
- Process repeats until all data is classified.



SPLITTING CRITERIA

Classification Trees:

- Uses Entropy & Information Gain
- Uses Gini Impurity (alternative to entropy)

Regression Trees:

- Uses Mean Squared Error (MSE)




FEATURE SELECTION IN DECISION TREES

Best Split? → Choose the feature with the highest Information Gain

Continuous Features? → Use threshold-based splits

SUMMARY

Key Takeaways

-  Decision Trees are easy to interpret but prone to overfitting
-  Random Forest reduces variance by combining multiple trees
-  Pruning helps avoid overfitting



THANK YOU

