SFFA

Random Forests and Applications in Financial Planning

SFFA

Brief Agenda

03. What is A Random Forest?

04. Ensemble Learning

05. Decision Trees

06. How it works ?

07. Terms and Hyperparameter

08. Dataset Preparation

09. Advantages & Disadvantages

10. Random Forest Financial Application

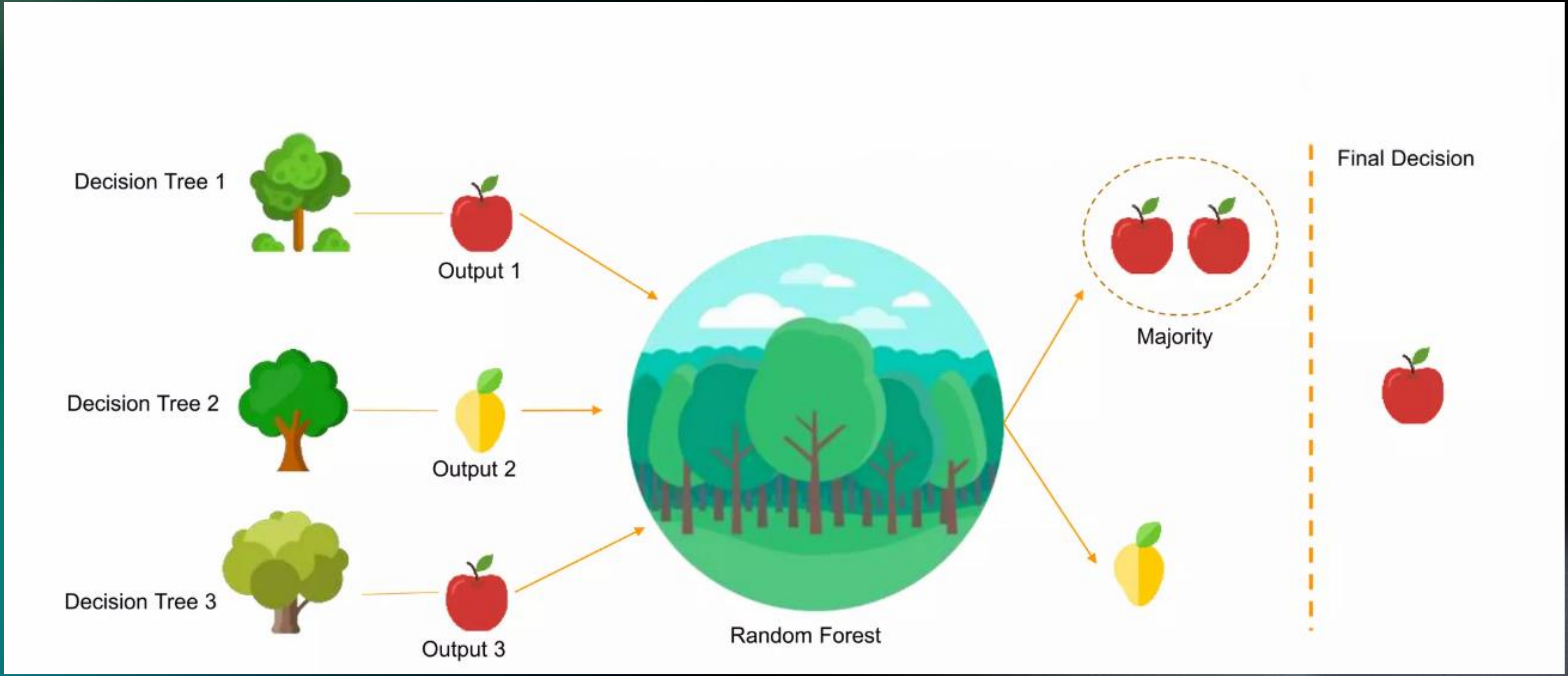
What is A Random Forest

Basic Concept

- A random forest algorithm is a supervised machine learning algorithm consisting of decision trees.
- It operates by constructing multiple Decision Trees during the training phase.
- The decision of the majority of the trees is chosen by the random forest as the final decision

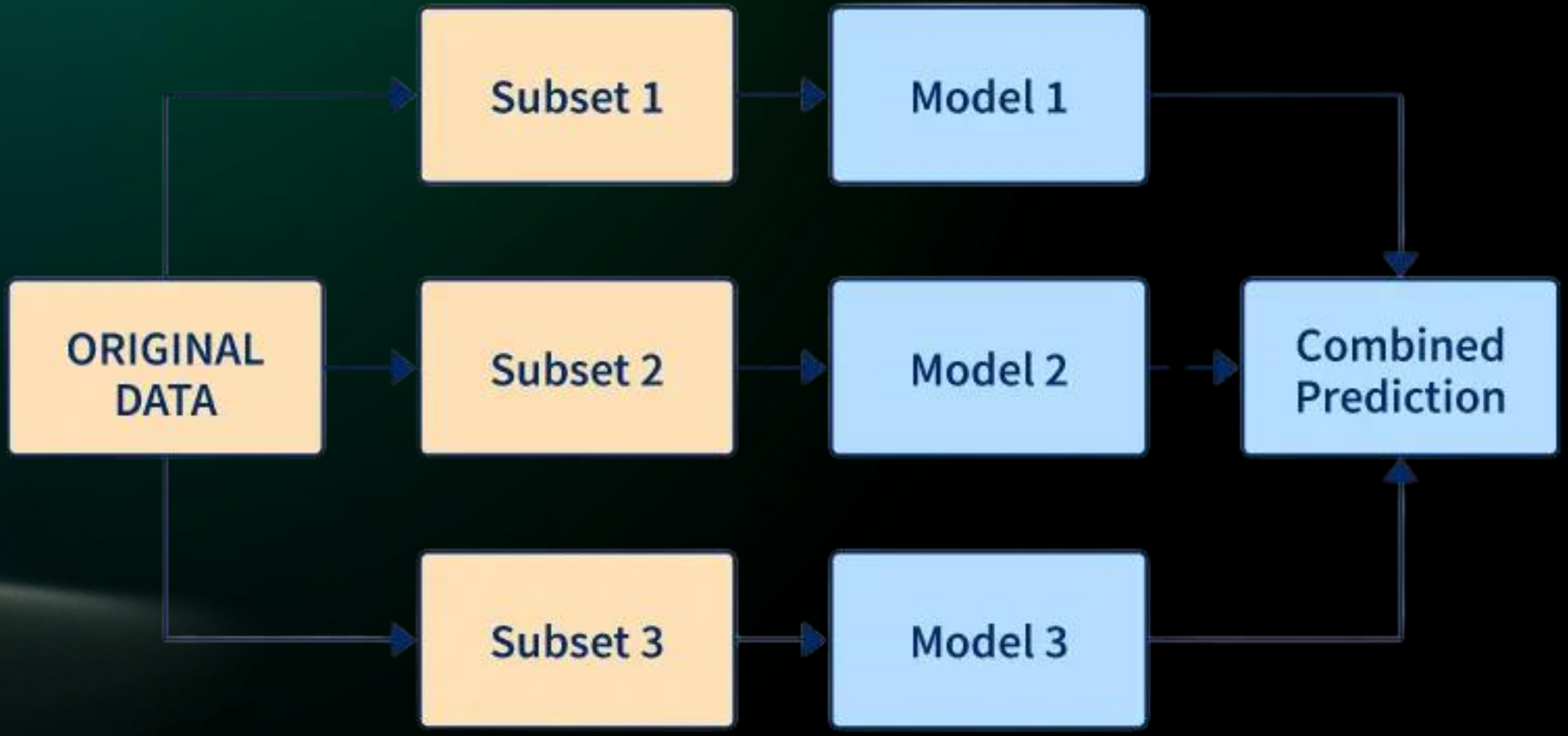
What is A Random Forest

Example



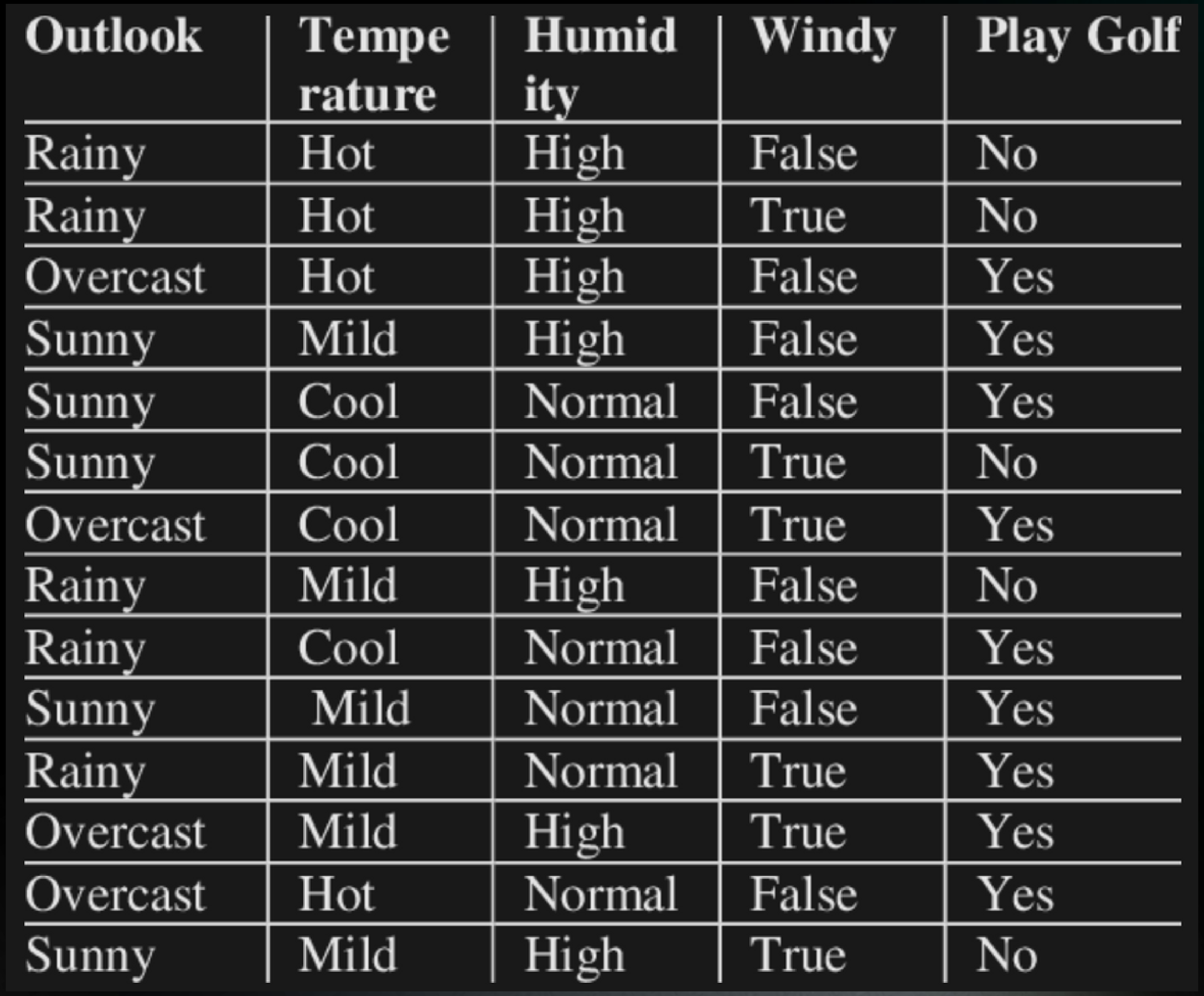
Ensemble Learning

Ensemble learning creates a stronger model by aggregating the predictions of multiple weak models. Random Forest is an example of ensemble learning where each model is a decision tree. The idea behind it is - the wisdom of the crowd. The majority vote aggregation can have better accuracy than the individual models.



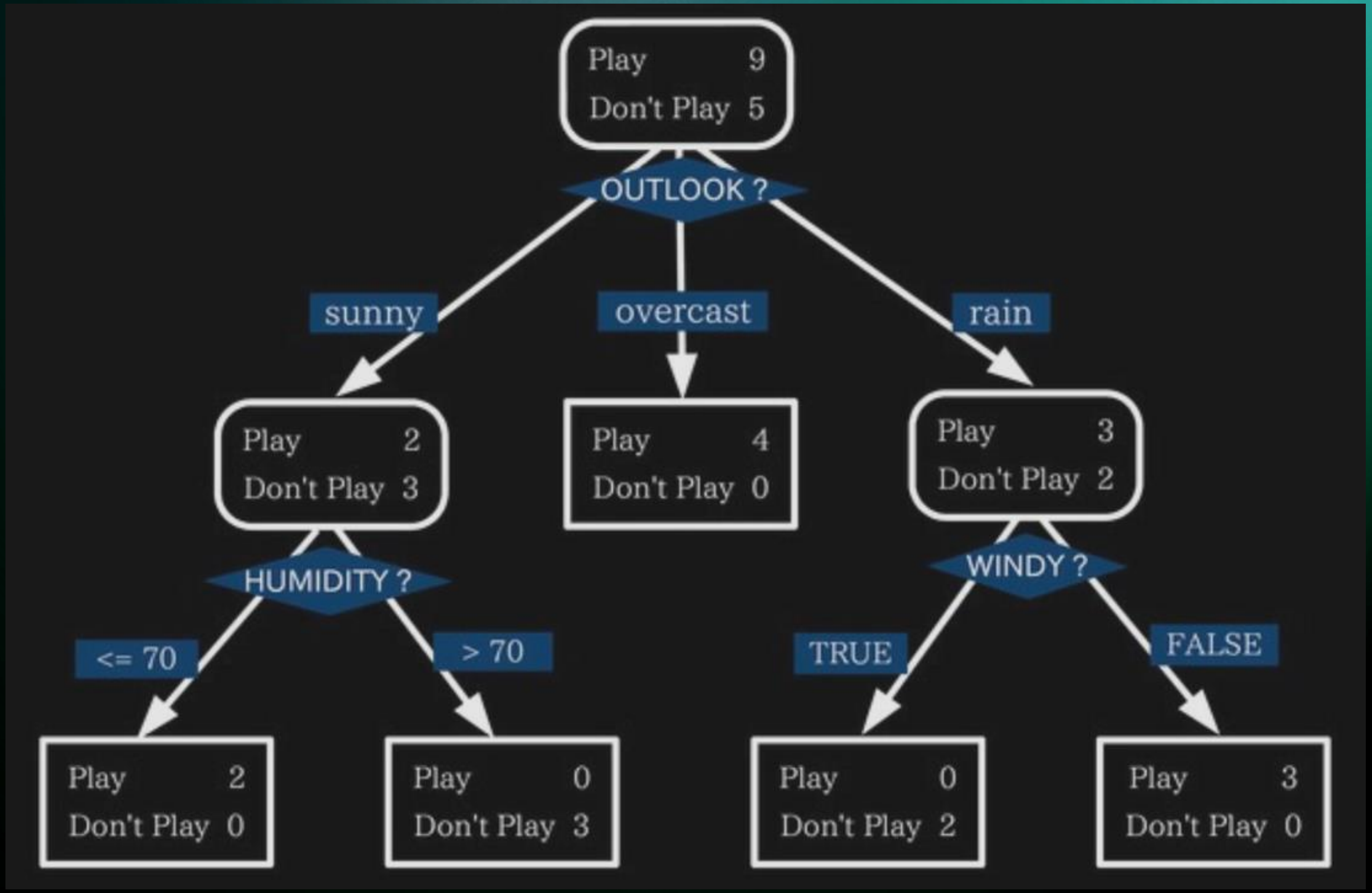
Decision Trees Example

Dataset



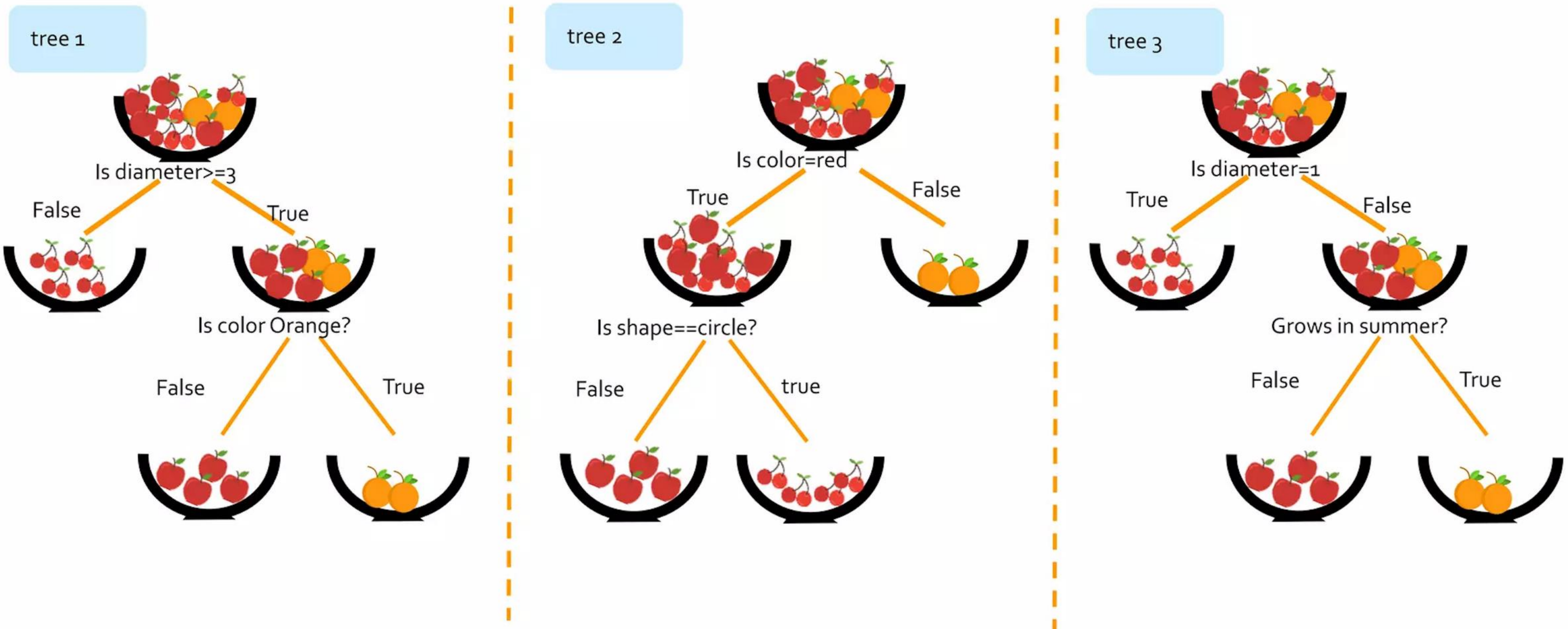
Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Desicion Tree

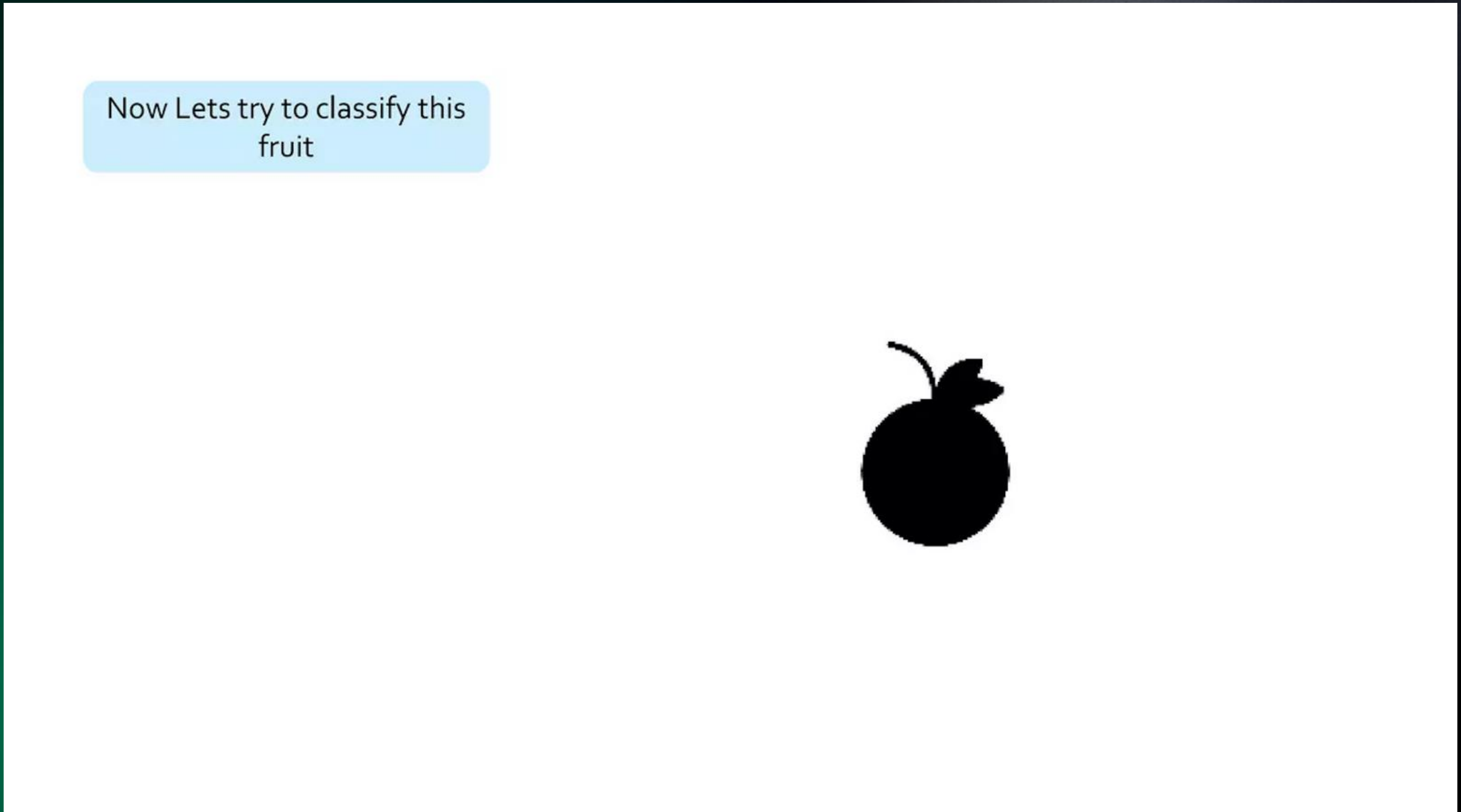


SFFA

How does a Random Forests Works?



How does a Random Forests Works?



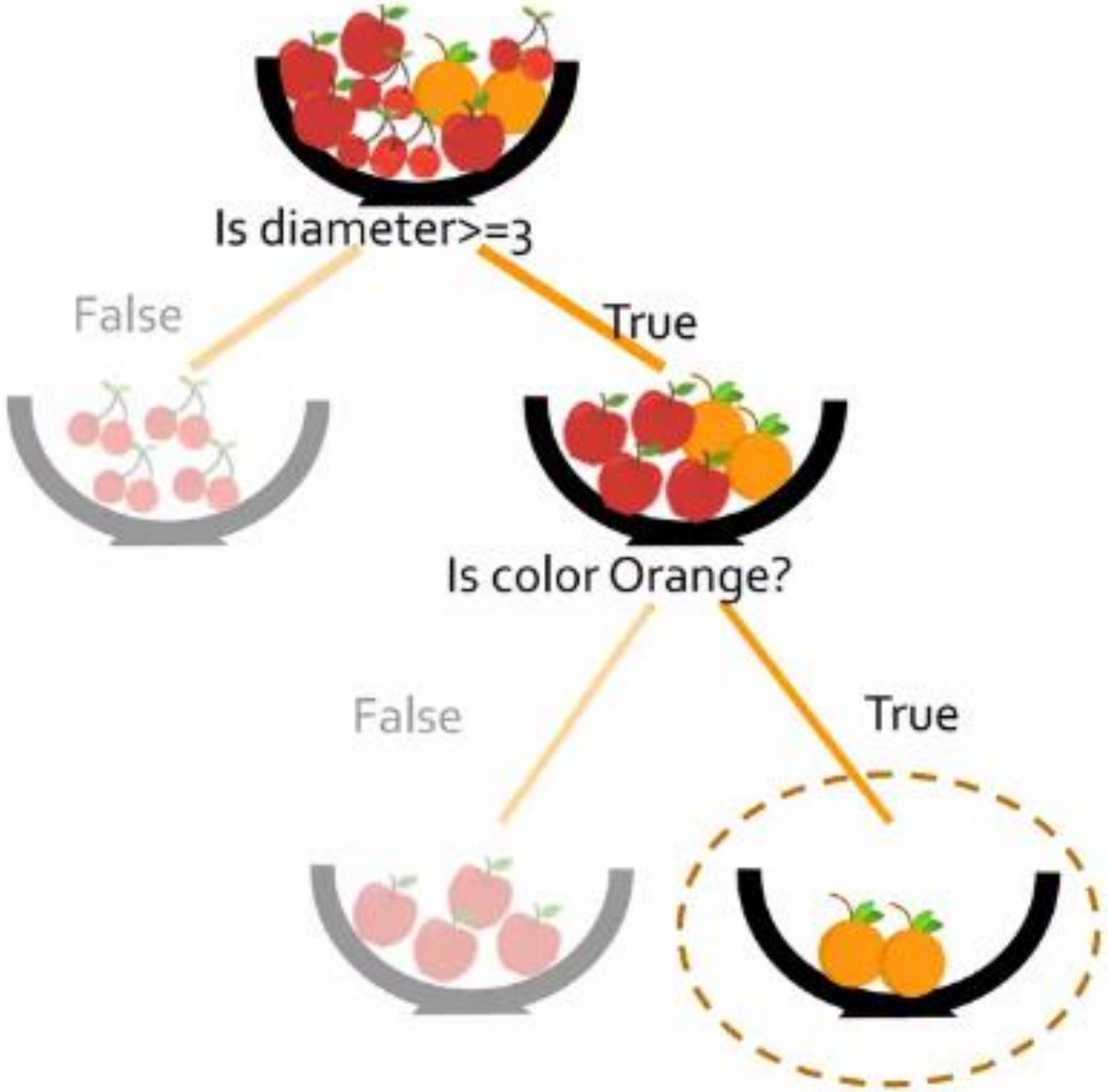
Now Lets try to classify this fruit

SFFA

How does a Random Forests Works?

Tree 1 classifies it as an orange

Diameter = 3
Colour = orange
Grows in summer = yes
SHAPE = CIRCLE

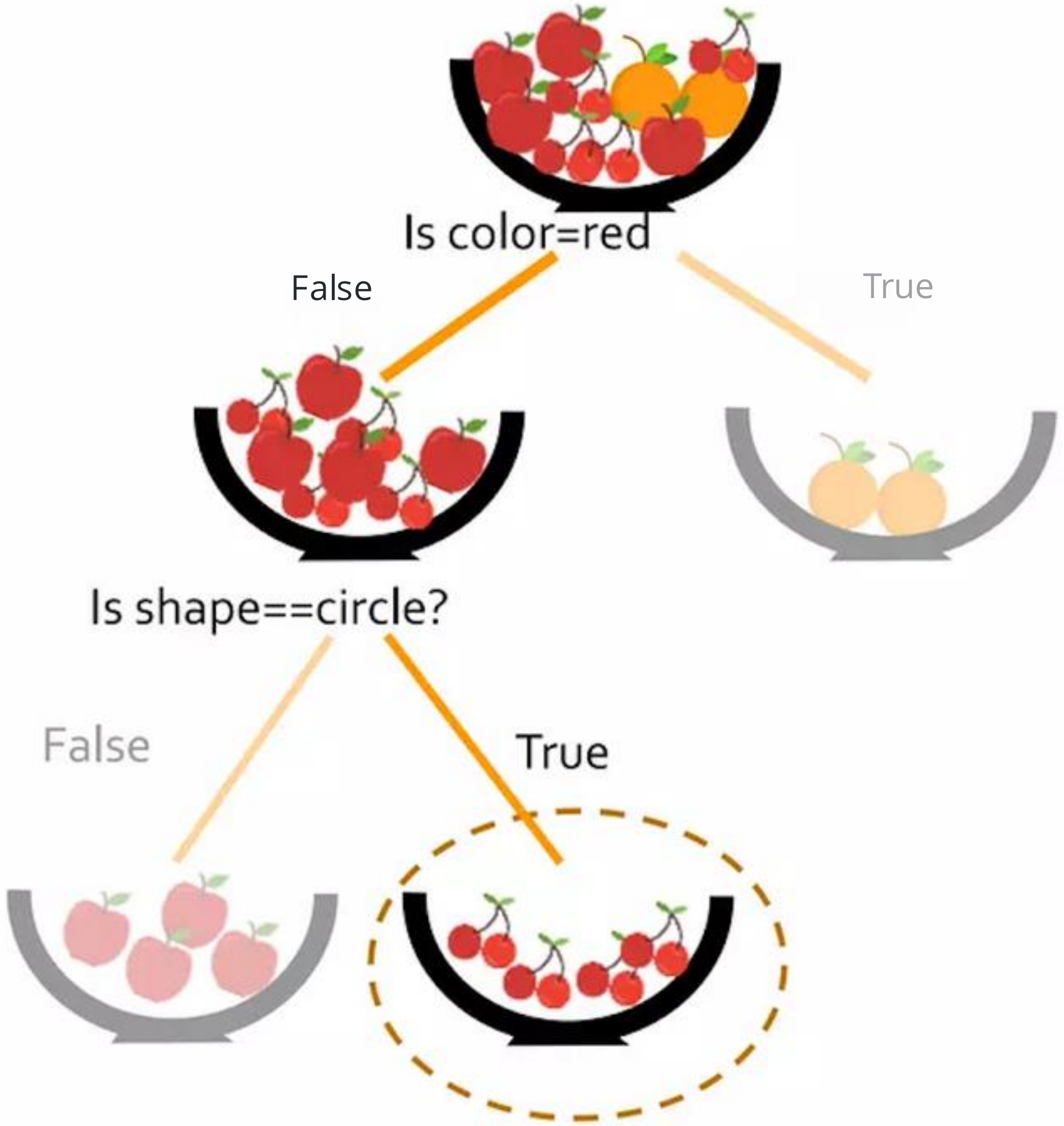


SFFA

How does a Random Forests Works?

Tree 2 classifies it as cherries

Diameter = 3
Colour = orange
Grows in summer = yes
SHAPE = CIRCLE

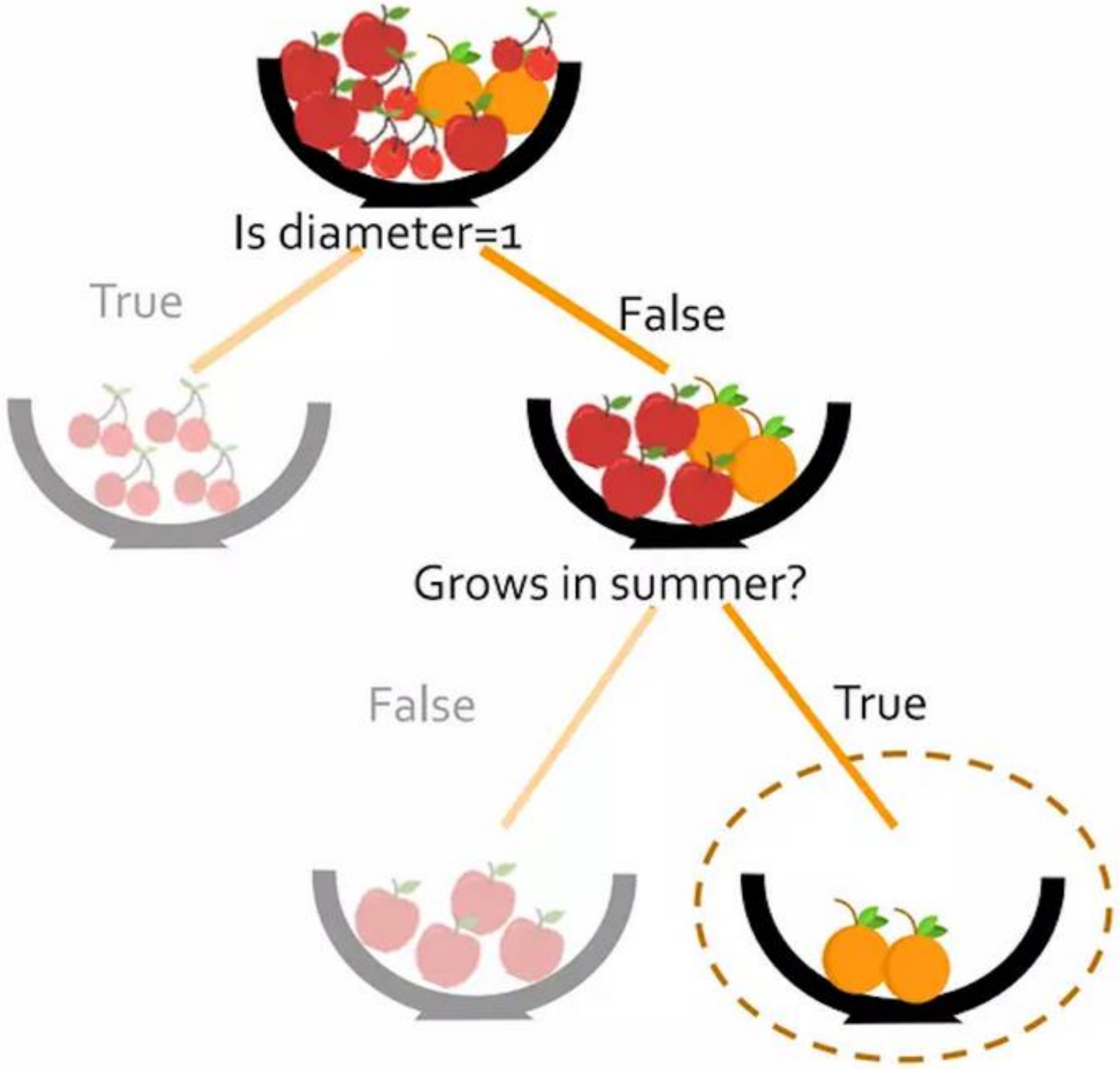


SFFA

How does a Random Forests Works?

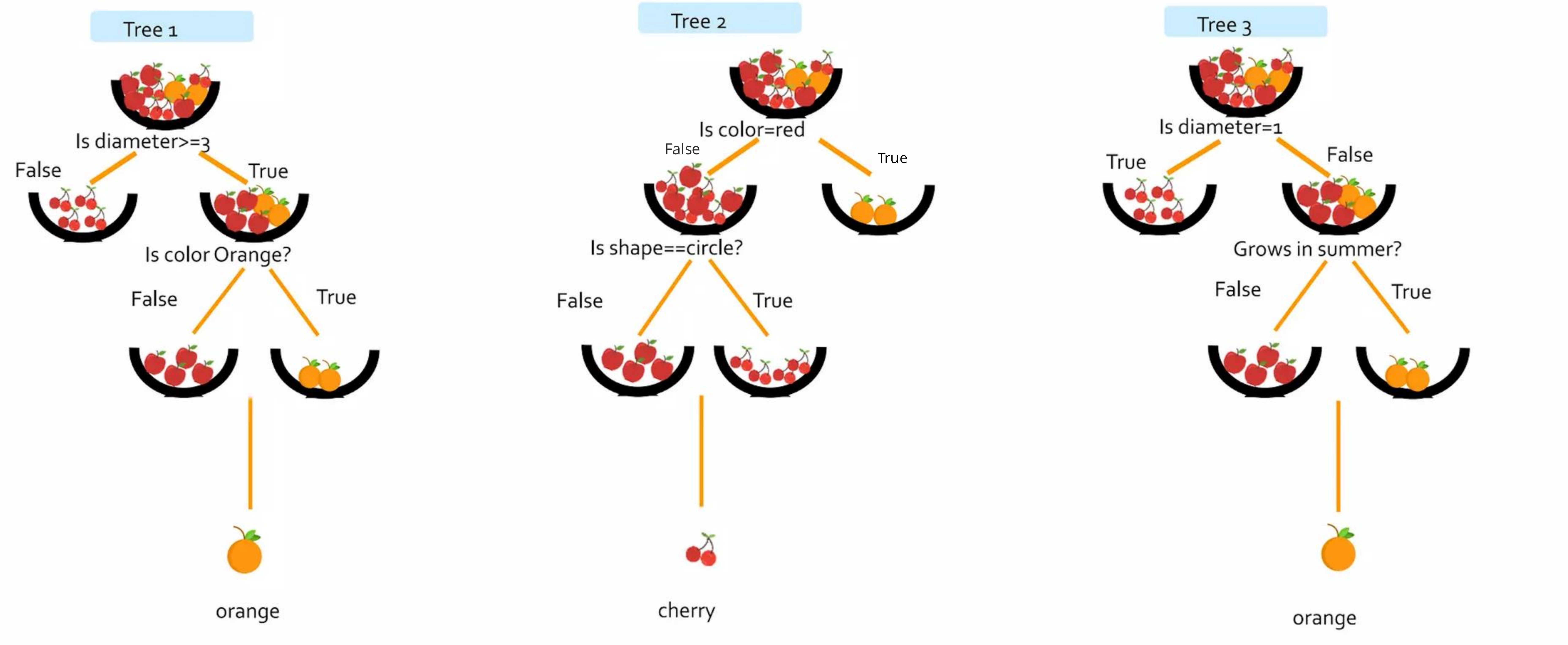
Tree 3 classifies it as orange

Diameter = 3
Colour = orange
Grows in summer = yes
SHAPE = CIRCLE



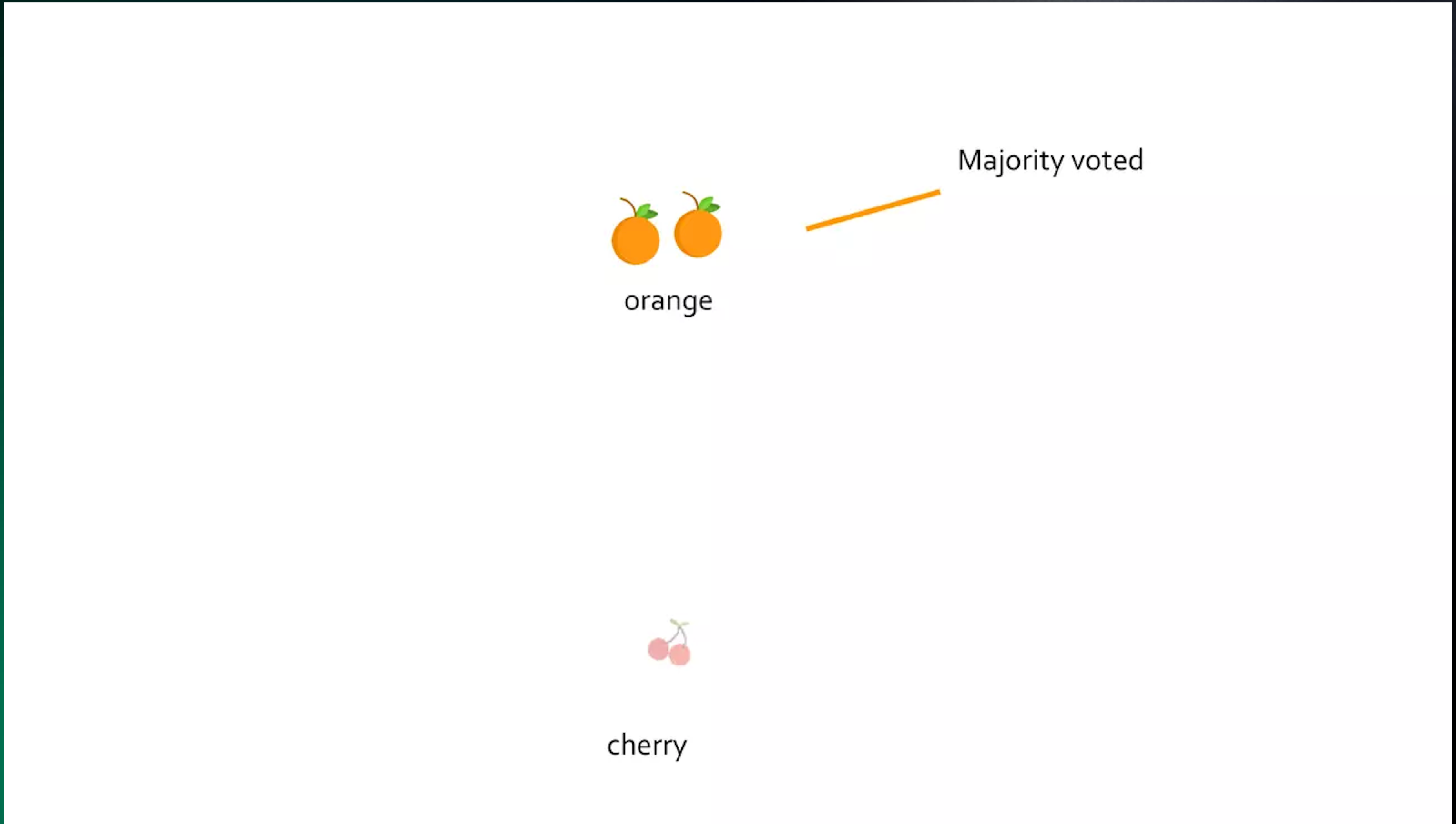
SFFA

How does a Random Forests Works?



SFFA

How does a Random Forests Works?



Important Terms

- Entropy: Is a measure of randomness or unpredictability in the data set.
- Information Gain: is a measure of the decrease in the entropy after the data set is split.
- Leaf Node: is a node that carries the classification or the decision.
- Decision Node: is a node that has two or more branches.
- Root Node: is the topmost decision node.

Important Hyperparameters

- Hyperparameters are used in random forests to either enhance the performance and predictive power of models or to make the model faster.

Predictive power hyperparameters

- n_estimators: Number of trees.
- mini_sample_leaf: Minimum number of leaves required to split an internal node.

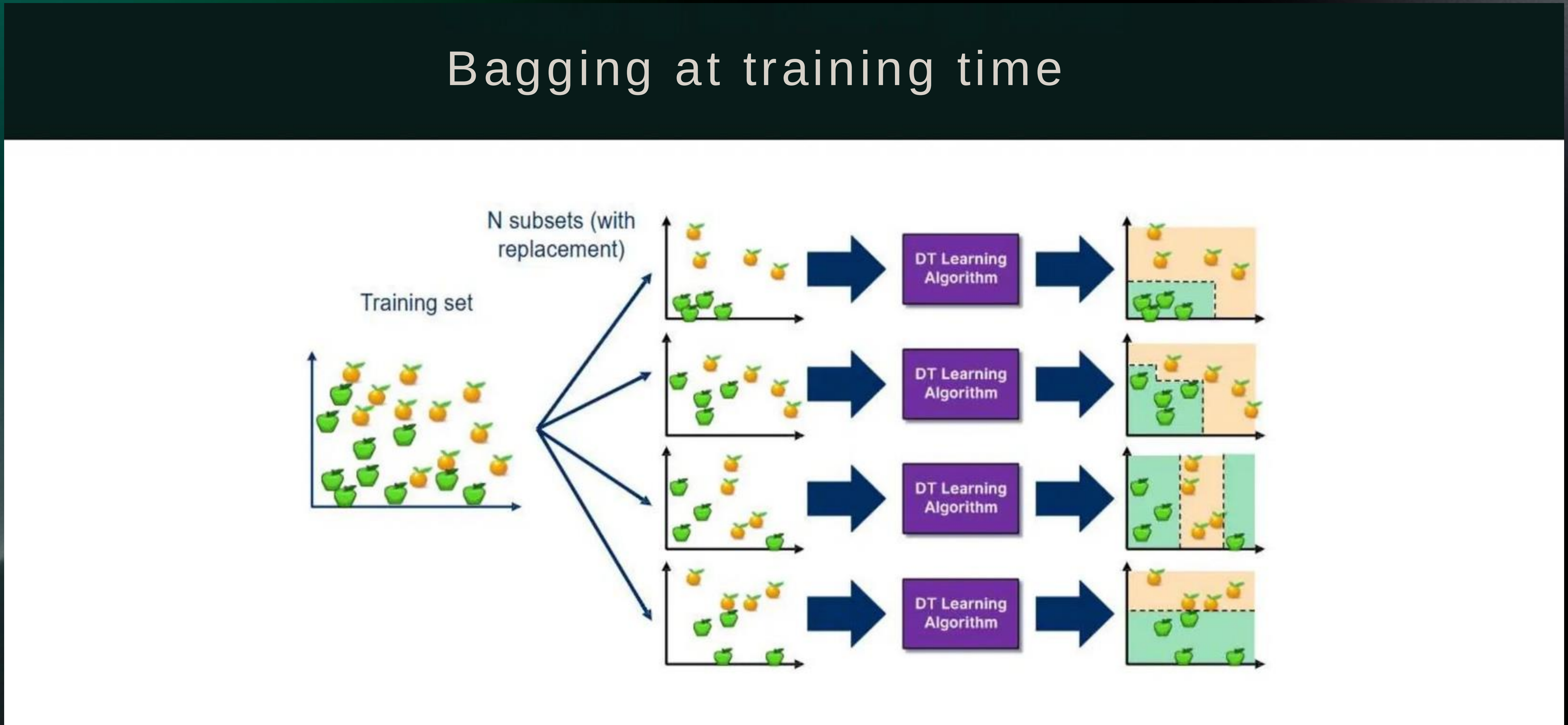
Speed hyperparameters

- n_jobs: Number processors allowed to use.
- random_state: Controls randomness of the sample.

Dataset Preparation

Bootstrap Aggregating (Bagging)

Creating a different training subset randomly from the original training dataset with replacement is called Bagging. With replacement refers to the bootstrap sample having duplicate elements. Reduces variance, helps to avoid overfitting.

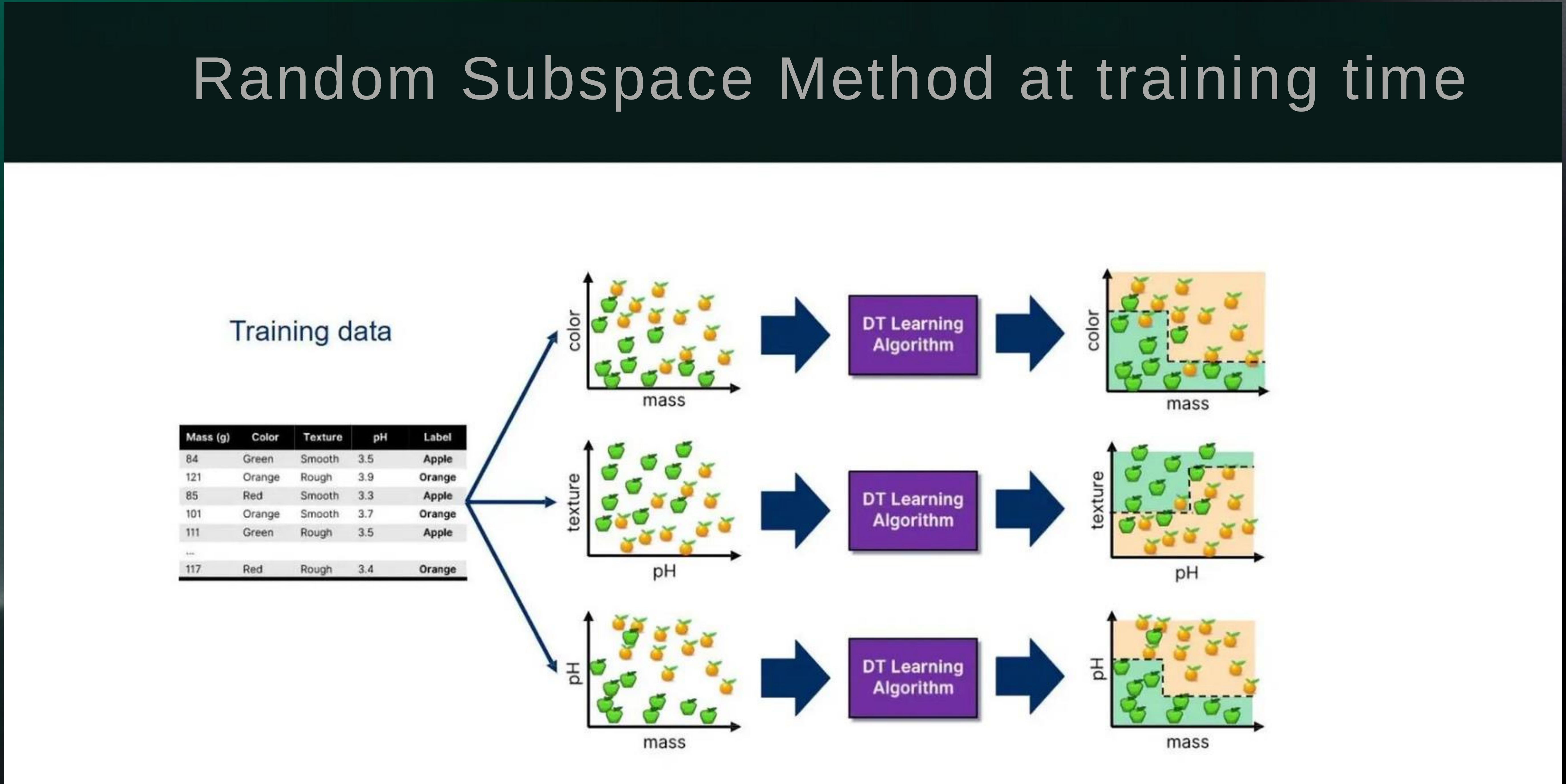


Bagging at training time

Dataset Preparation

Random Subspace Method(Feature Bagging)

While building the tree, for splitting, a randomly selected subset of the features are used. So, the trees are more different having low correlation.



Random Subspace Method at training time

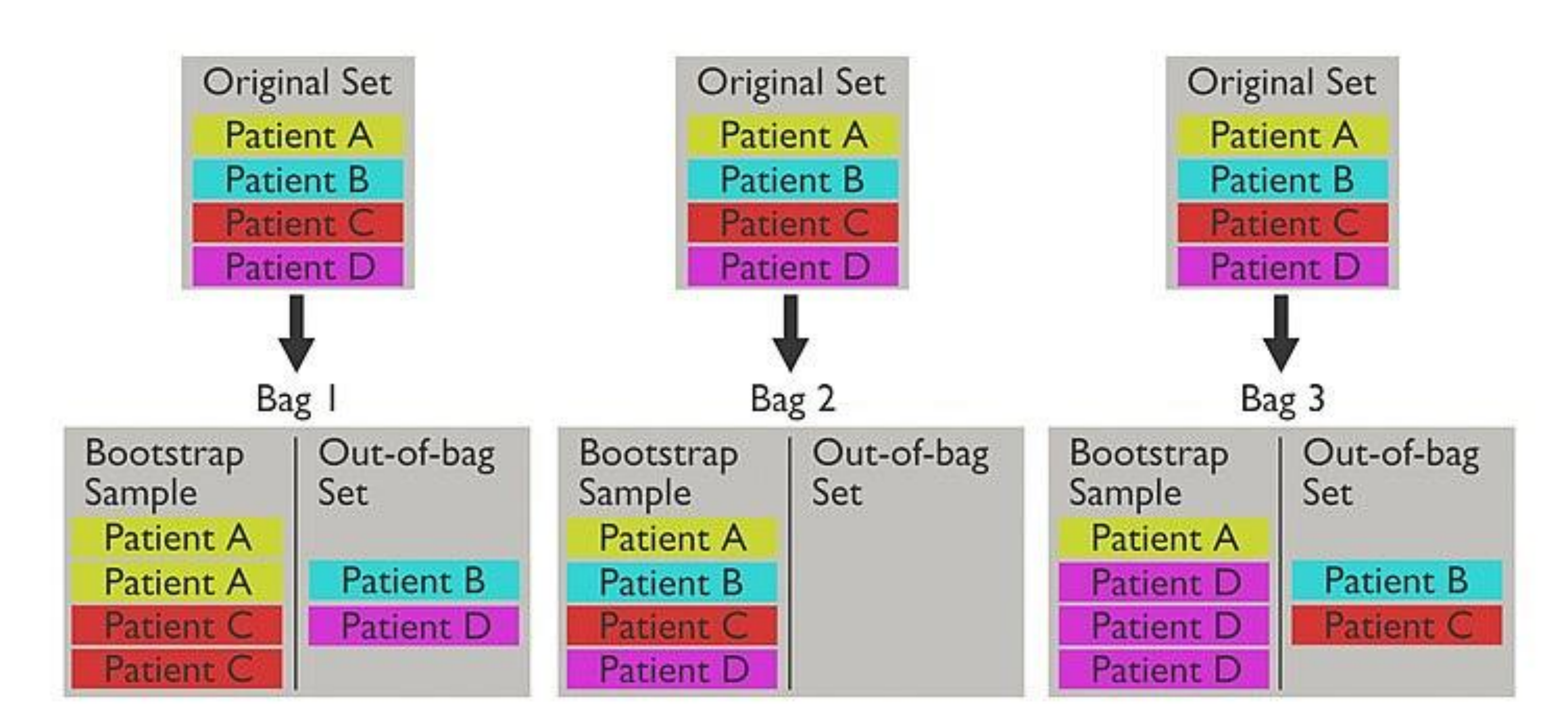
Training data

Mass (g)	Color	Texture	pH	Label
84	Green	Smooth	3.5	**Apple**
121	Orange	Rough	3.9	**Orange**
85	Red	Smooth	3.3	**Apple**
101	Orange	Smooth	3.7	**Orange**
111	Green	Rough	3.5	**Apple**
...				
117	Red	Rough	3.4	**Orange**

Dataset Preparation

Out-Of-Bag (OOB)

The out-of-bag dataset represents the remaining elements which were not in the bootstrap dataset (Unseen). Used for cross validation or to evaluate the performance.



Original Set
Patient A
Patient B
Patient C
Patient D

Bag 1

Bootstrap Sample
Patient A
Patient A
Patient C
Patient C

Out-of-bag Set
Patient B
Patient D

Original Set
Patient A
Patient B
Patient C
Patient D

Bag 2

Bootstrap Sample
Patient A
Patient B
Patient C
Patient D

Out-of-bag Set

Original Set
Patient A
Patient B
Patient C
Patient D

Bag 3

Bootstrap Sample
Patient A
Patient D
Patient D
Patient D

Out-of-bag Set
Patient B
Patient C

Advantages and Disadvantages of Random Forests

Advantages

- Versatile uses.
- Easy-to-understand hyperparameters.
- Classifier doesn't overfit with enough trees.

Disadvantages

- Increased accuracy requires more trees.
- More trees slow down model.
- Can't describe relationships within data.

Applications of Random Forest

- Health Care: Health professionals use random forest systems to diagnose patients.
- Stock Market: Financial analysts use it to identify potential markets for stocks.
- E-Commerce: Through this system, e-commerce vendors can predict the preferences of customers based on past consumption behavior.

Why Random Forests is a Good Fit for Financial Application

- Handles Complex Relationships: Financial data often has non-linear relationships (e.g., income level, spending habits, and risk tolerance).
- Robustness to Noise and Outliers: Financial data is typically noisy and may contain outliers, such as extremely high or low incomes or unusual spending habits
- Works Well with Categorical and Continuous Variables: Financial data often contains both categorical (e.g., marital status, financial goals) and continuous variables (e.g., income, expenses).

How we will apply Random Forest

Objective

Predict the most suitable budgeting rule (50/30/20, 70/20/10, and 60/20/20) for users based on financial profiles.

Why Choose Bagging?

- Reduces variance and prevents overfitting.
- Enhances prediction accuracy and stability.
- Utilizes Out-of-Bag (OOB) evaluation for internal validation.

Data Collection

- Throw open data platform.
- By collecting surveys.

Thank You

Any question ?