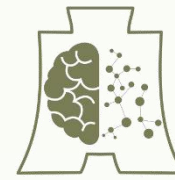


# Random Forest Introduction

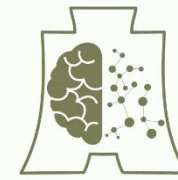
National Kaohsiung University of Sciences and Technology  
Department of Finance and Information, Professor  
AI Fintech Center, Director  
Lin, Ping-Chen

- 什麼是隨機森林
- 隨機森林演算法
- 隨機森林的優點
- 隨機森林的缺點



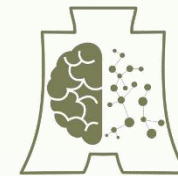
# Random Forest Introduction

## 什麼是隨機森林?



- 隨機森林是一種監督學習演算法，是由Leo Breiman和Adele Cutler商標註冊的常用機器學習演算法。
- 結合多個決策樹的輸出來達成單一結果
- 可以處理分類和迴歸問題
- 最靈活且易於使用的演算法之一。在給定的數據樣本上創建決策樹，從每個樹中獲取預測，並通過投票選擇最佳解決方案。
- 是一個相當不錯的特徵重要性指標
- 隨機森林演算法結合多個決策樹，形成一片樹林，因此得名隨機森林
- 在隨機森林分類器中，森林中的樹木數量越多，準確性越高

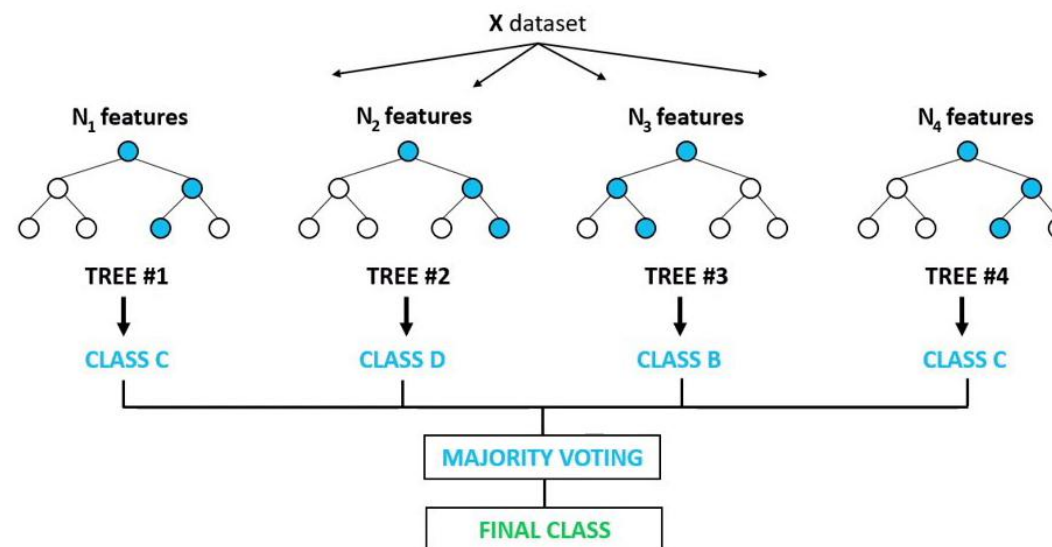
## 隨機森林演算法



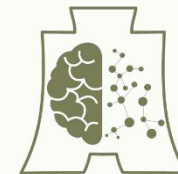
- 隨機森林演算法直觀分為兩個階段
- 從總共 $m$ 個特徵中選擇 $k$ 個特徵並構建隨機森林
  1. 隨機選擇 $k$ 個特徵，其中 $k < m$
  2. 在這 $k$ 個特徵中，使用最佳分割點計算節點 $d$ 。
  3. 使用最佳分割將節點分割成子節點。
  4. 重複步驟1到3，直到達到 $l$ 個節點
  5. 重複步驟1到4  $n$ 次，以創建 $n$ 個樹木，從而構建森林

- 第二階段使用訓練好的隨機森林演算法進行預測
  1. 將測試特徵帶入每棵隨機創建的決策樹，預測結果並存儲預測結果
  2. 計算每個預測目標的投票數
  3. 將投票數最高的預測目標作為隨機森林演算法的最終預測

### Random Forest Classifier



## 隨機森林的優點

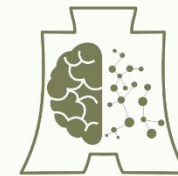


AI.FINTECH

AI 金融 科技 中心

- 隨機森林演算法可用於解決分類和迴歸問題
- 被認為是非常準確和健壯的模型，因為它使用大量的決策樹進行預測
- 隨機森林採用所有決策樹的預測平均值來抵消偏差，因此不會遇到過擬合問題
- 隨機森林分類器可以處理缺失值，有兩種方法
  - 使用中位數替換連續變量
  - 計算缺失值的鄰近加權平均值
- 隨機森林分類器可用於特徵選擇，從訓練數據集中選擇最重要的特徵

## 隨機森林的缺點



AI.FINTECH

AI 金融 科技 中心

- 隨機森林的最大缺點是計算複雜度
- 因為使用大量的決策樹進行預測，隨機森林在做預測時非常慢
- 森林中的所有樹木必須對相同的輸入進行預測，然後進行投票，因此這是一個耗時的過程
- 與決策樹相比，模型難以解釋。



- <https://www.ibm.com/topics/random-forest>
- <https://www.kaggle.com/code/prashant111/random-forest-classifier-tutorial>
- <https://builtin.com/data-science/random-forest-algorithm>



**Thank you.**

