# Enhancing Predictive Analytics in Sports Using Random Forests

By: Johnny Rosas

## Objective:

- Enhance predictive analytics in the NBA using the Random Forest algorithm

## Goal:

- Improve strategic decision-making for team composition and game strategies.

# Components:

- Data Collection
- Data Processing
- Features
- Model Training w/ Random Forest
- Command-Line Interface(WSL)

# Random Forest Algorithm (RFA)

**Definition -** is a learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees.

**Advantages:**

- Reduces overfitting and improves accuracy
- Provides insights into value significance
- Can handle large data sets with big dimensionality

# How it works (RFA)

## Step 1: Bootstrap Sampling

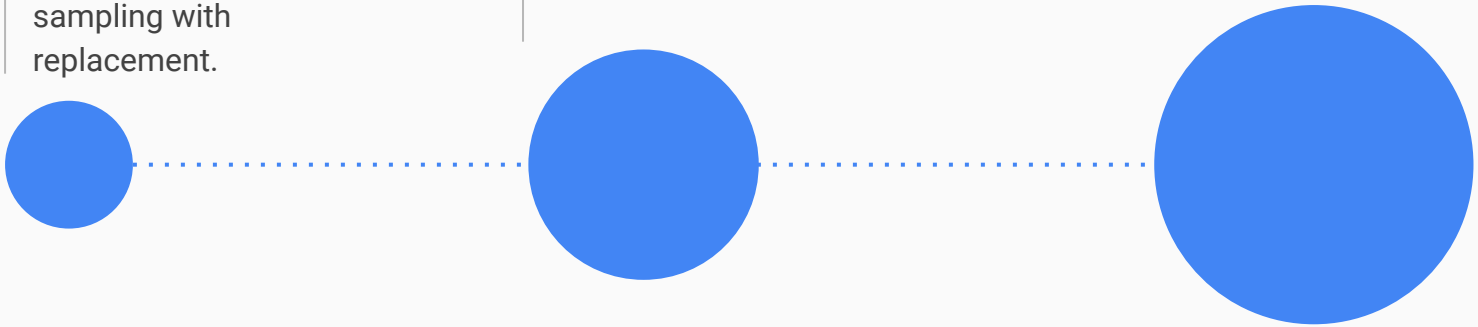
Create multiple subsets of the training data by sampling with replacement.

## Step 2: Decision Trees

Train a decision tree on each subset.

## Step 3: Combine

Combine the prediction of each tree to form the final prediction.



# Program Overview:

Data Collection:

- Fetch teams, rosteres, and game data from ESPN

Data Processing:

- Convert dates, extract scores, determine home/away status

Features:

- Create features like points scored, points allowed, home/away status

Model Training:

- Train Random Forest using processed data
- Evaluate model performance using accuracy, precision, recall, and F-1 score

Command-Line Interface (WSL):

- Uses command line to fetch and display data and model predictions

README

- Use read me to help you setup the program and learn how to use it