1 .RANDOM FOREST

-Thuật toán Random Forest là một kỹ thuật học cây mạnh mẽ trong Machine Learning . Nó hoạt động bằng cách tạo ra một số Cây quyết định trong giai đoạn đào tạo. Mỗi cây được xây dựng bằng cách sử dụng một tập hợp con ngẫu nhiên của tập dữ liệu để đo lường một tập hợp con ngẫu nhiên các tính năng trong mỗi phân vùng. Tính ngẫu nhiên này tạo ra sự biến đổi giữa các cây riêng lẻ, giảm nguy cơ trang bị quá mức và cải thiện hiệu suất dự đoán tổng thể.

-Trong dự đoán, thuật toán tổng hợp kết quả của tất cả các cây, bằng cách bỏ phiếu (đối với nhiệm vụ phân loại) hoặc bằng cách lấy trung bình (đối với nhiệm vụ hồi quy). Quá trình ra quyết định hợp tác này, được hỗ trợ bởi nhiều cây với thông tin chi tiết của chúng, cung cấp một ví dụ về kết quả ổn định và chính xác .

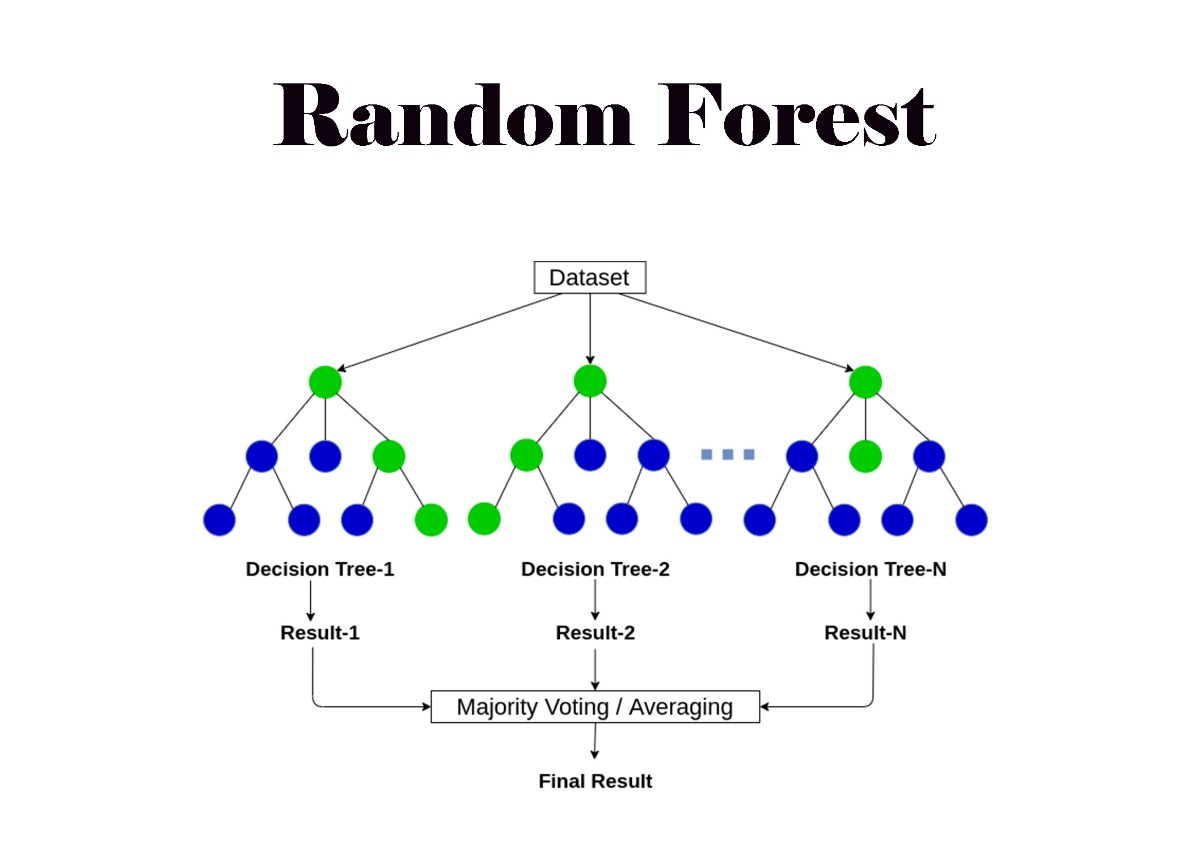
-Random Forest được sử dụng rộng rãi cho các chức năng phân loại và hồi quy, được biết đến với khả năng xử lý dữ liệu phức tạp, giảm tình trạng trang bị quá mức và cung cấp dự báo đáng tin cậy trong các môi trường khác nhau.

A diagram of a training data

Description automatically generated

2. Nguyên lí hoạt động

- Tập hợp các Cây quyết định: Random Forest tận dụng sức mạnh của việc học tổng hợp bằng cách xây dựng một tập hợp Cây quyết định . Những cây này giống như các chuyên gia riêng lẻ, mỗi chuyên gia chuyên về một khía cạnh cụ thể của dữ liệu. Điều quan trọng là chúng hoạt động độc lập, giảm thiểu rủi ro mô hình bị ảnh hưởng quá mức bởi các sắc thái của một cây duy nhất.



- Lựa chọn tính năng ngẫu nhiên: Để đảm bảo rằng mỗi cây quyết định trong tập hợp mang lại một góc nhìn độc đáo, Random Forest sử dụng lựa chọn tính năng ngẫu nhiên . Trong quá trình huấn luyện mỗi cây, một tập hợp con các đặc tính ngẫu nhiên sẽ được chọn. Tính ngẫu nhiên này đảm bảo rằng mỗi cây tập trung vào các khía cạnh khác nhau của dữ liệu, thúc đẩy một tập hợp các yếu tố dự đoán đa dạng trong quần thể.

- Tổng hợp hoặc đóng bao Bootstrap: Kỹ thuật đóng bao là nền tảng trong chiến lược đào tạo của Random Forest, bao gồm việc tạo nhiều mẫu bootstrap từ tập dữ liệu gốc, cho phép lấy mẫu các phiên bản thay thế. Điều này dẫn đến các tập hợp dữ liệu khác nhau cho mỗi cây quyết định, tạo ra sự thay đổi trong quá trình đào tạo và làm cho mô hình trở nên mạnh mẽ hơn.

- Ra quyết định và biểu quyết: Khi đưa ra dự đoán, mỗi cây quyết định trong Random Forest sẽ bỏ phiếu. Đối với các nhiệm vụ phân loại , dự đoán cuối cùng được xác định theo chế độ (dự đoán thường xuyên nhất) trên tất cả các cây. Trong các tác vụ hồi quy , giá trị trung bình của các dự đoán cây riêng lẻ được lấy. Cơ chế bỏ phiếu nội bộ này đảm bảo quá trình ra quyết định tập thể và cân bằng.

3. Đặc điểm

+ Bản chất Ensemble: Random Forest kết hợp nhiều cây quyết định với nhau. Mỗi cây quyết định được huấn luyện trên một phần của dữ liệu (bootstrap sample), và một số lượng ngẫu nhiên các đặc trưng (feature) được chọn cho mỗi nút trong quá trình phân chia. Việc kết hợp nhiều cây giúp giảm thiểu overfitting và cải thiện tính tổng quát hóa của mô hình.

+ Tính linh hoạt: Random Forest có thể được áp dụng cho cả bài toán phân loại và hồi quy. Đối với bài toán phân loại, nó có thể xử lý nhiều loại dữ liệu và có khả năng xử lý cả dữ liệu không cân bằng.

+ Hiệu suất cao: Random Forest thường cho hiệu suất tốt trên nhiều loại dữ liệu, bao gồm cả dữ liệu có tính không tuyến tính và dữ liệu không cân bằng. Nó thường không đòi hỏi nhiều thời gian để điều chỉnh các siêu tham số và thường đưa ra các dự đoán tốt mà không cần quá nhiều tinh chỉnh.

+ Khả năng xử lý dữ liệu lớn: Random Forest có thể xử lý được các tập dữ liệu lớn với số lượng đặc trưng lớn mà vẫn giữ được hiệu suất.

+ Interpretability (Khả năng giải thích): Mặc dù mỗi cây quyết định trong Random Forest có thể không dễ hiểu, nhưng khi kết hợp lại với nhau, mô hình có thể cung cấp cái nhìn tổng quan về cách các đặc trưng ảnh hưởng đến dự đoán cuối cùng.

**2 Đánh giá mô hìnA graph with numbers and a number of blue squares

Description automatically generated with medium confidenceh**

A screenshot of a computer program

Description automatically generated