**Random Forest (Rừng Ngẫu Nhiên)** là một mô hình học máy thuộc loại tập hợp cây quyết định (tree ensemble). Nó được xây dựng dựa trên phương pháp Bagging (Bootstrap Aggregating) và sử dụng lấy mẫu với thay thế (sampling with replacement) để tạo ra nhiều cây quyết định độc lập.

Dưới đây là cách Random Forest hoạt động:

1. **Lấy Mẫu với Thay Thế:**
   * Cho một tập dữ liệu huấn luyện có kích thước �*N*, Random Forest tạo ra �*B* tập dữ liệu con bằng cách lấy mẫu �*N* mẫu từ tập dữ liệu gốc với thay thế. Điều này có nghĩa là một mẫu có thể xuất hiện nhiều lần trong một tập dữ liệu con nhưng không xuất hiện trong tập dữ liệu con khác.
2. **Xây Dựng Cây Quyết Định:**
   * Đối với mỗi tập dữ liệu con, một cây quyết định (Decision Tree) được xây dựng. Mỗi cây được xây dựng bằng cách chọn biến tại mỗi nút dựa trên Information Gain, nhưng chỉ trên một số lượng ngẫu nhiên �*m* biến (số biến nhỏ hơn tổng số biến). Điều này tạo ra sự đa dạng giữa các cây.
3. **Bình Chọn (Classification) hoặc Trung Bình (Regression):**
   * Đối với bài toán phân loại, kết quả của Random Forest là kết quả của việc bình chọn của tất cả cây. Đối với bài toán hồi quy, kết quả là giá trị trung bình của tất cả cây.

**Ưu Điểm của Random Forest:**

* **Đa Dạng và Ổn Định:** Do sử dụng lấy mẫu với thay thế và xây dựng các cây dựa trên số lượng ngẫu nhiên của các biến, Random Forest tạo ra một tập hợp đa dạng các cây, giúp ổn định và giảm overfitting.
* **Hiệu Suất Cao:** Thường có hiệu suất tốt trên nhiều loại dữ liệu và không yêu cầu nhiều siêu tham số được điều chỉnh.

**Nhược Điểm của Random Forest:**

* **Khó Để Giải Thích:** Do sự phức tạp của tập hợp cây, Random Forest có thể khó giải thích so với một cây quyết định đơn lẻ.
* **Tăng Chi Phí Tính Toán:** So với một cây quyết định đơn lẻ, Random Forest có thể tăng chi phí tính toán vì cần xây dựng và đánh giá nhiều cây.