

EXPERIMENT - 6

**Aim:** Classification modeling

1. Choose classifier for classification problem.
2. Evaluate the performance of classifier.

# Theory:

## Introduction to Classification

Classification is a type of supervised learning where the goal is to categorize data into predefined labels or classes. It involves training a model on labeled data so that it can make predictions on new, unseen data. In this experiment, we use the **Random Forest Classifier** to predict hair fall based on different features.

## Dataset and Preprocessing Feature Selection:

* + The dataset consists of various attributes related to hair fall.
  + The target variable (hair\_fall) is a categorical feature indicating whether hair fall occurs.
  + The independent variables (X) consist of different factors that may contribute to

hair fall.

## Data Splitting:

* + The dataset is split into **training (80%)** and **testing (20%)** sets using train\_test\_split().
  + The training set is used to train the model, while the testing set evaluates its performance.

## Feature Scaling:

* + Since features may have different units and magnitudes, we apply **Standardization**

using StandardScaler().

* + Standardization transforms the features to have a **mean of 0** and **standard deviation of 1**, improving the model's performance.

## Random Forest Classifier

Random Forest is an **ensemble learning method** that constructs multiple decision trees and combines their outputs for more accurate predictions.

## Key Advantages of Random Forest:

* Handles missing values and noise well.
* Reduces overfitting compared to a single decision tree.
* Can handle both classification and regression tasks.

## Working of Random Forest:

* 1. Multiple decision trees are trained on different random subsets of the training data.
  2. Each tree makes a prediction, and the majority vote is taken for classification.
  3. The final prediction is based on the combined outputs of all trees.

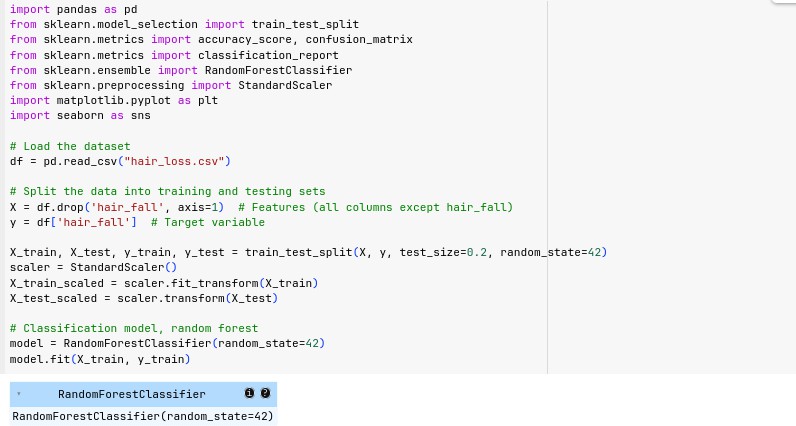
## Model Evaluation Accuracy Score:

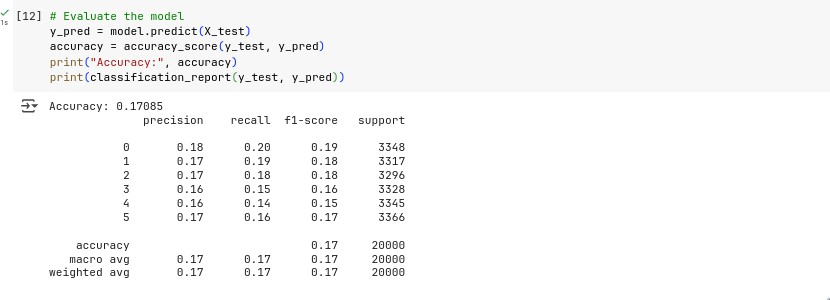
* + The **accuracy score** is calculated using accuracy\_score(y\_test, y\_pred).
  + It measures how many predictions were correct compared to the total number of samples.

## Classification Report:

* + Provides metrics like **Precision, Recall, and F1-score** for each class.
  + classification\_report(y\_test, y\_pred) generates the report.

# Program:

****



****

**Conclusion:** Thus, we have successfully implemented classification modeling using random forest