

# Ensemble Techniques And Its Types-5

## Assignment Questions



**Q1.** You are working on a machine learning project where you have a dataset containing numerical and categorical features. You have identified that some of the features are highly correlated and there are missing values in some of the columns. You want to build a pipeline that automates the feature engineering process and handles the missing values.

**Design a pipeline that includes the following steps:**

- Use an automated feature selection method to identify the important features in the dataset.
- Create a numerical pipeline that includes the following steps:
  - Impute the missing values in the numerical columns using the mean of the column values.
  - Scale the numerical columns using standardisation.
- Create a categorical pipeline that includes the following steps:
  - Impute the missing values in the categorical columns using the most frequent value of the column.
  - One-hot encode the categorical columns.
- Combine the numerical and categorical pipelines using a ColumnTransformer.
- Use a Random Forest Classifier to build the final model.
- Evaluate the accuracy of the model on the test dataset.

**Note:** Your solution should include code snippets for each step of the pipeline, and a brief explanation of each step. You should also provide an interpretation of the results and suggest possible improvements for the pipeline.

**Q2.** Build a pipeline that includes a random forest classifier and a logistic regression classifier, and then use a voting classifier to combine their predictions. Train the pipeline on the iris dataset and evaluate its accuracy.

**Note:** Create your assignment in Jupyter notebook and upload it to GitHub & share that github repository link through your dashboard. Make sure the repository is public.