MACHINE LEARNING : Cross Validation (KNN, DT & RF algorithm's)

Breast cancer data includes 569 observations of cancer biopsies, each with 32 characteristics (variables). One feature is an identification number (ID), another is a cancer diagnosis, and 30 are numerical laboratory measurements. The diagnosis is coded as "M" to indicate malignant or "B" for indicate benign.

You were hired to develop a classification algorithm to predict the diagnostic. Your goal is to apply cross-validation first to identify the better algorithm (accuracy) to be used (at least consider Decision Trees, Random Forest and KNN). After the identification you should develop a strategy to improve the accuracy based on hypermeters to achieve the best % of accuracy.

Note: Many techniques will be considered to score your work, including : EDA, Statistical analysis, outlier's management, normalization, feature engineering, duplicate lines, insights descriptions based on your own charts (matplotlib, seaborn, ...) and train-test split datasets.

Your Exercise deliverable should be : **Original dataset to be used + 1 Jupyter Notebook**

Submit a zip file with both on your course website.