Challenge 3 (Summary Report)

Name: **Anwai Archit**

Date of Submission**: 15th April 2021**

Subject: **Machine and Deep Learning (Medical Imaging and Applications)**

Requisite Points of Concern:

1. Dataset – Optical Recognition of Handwritten Digits.
2. Using Scikit-Learn Library.
3. Training and Testing the Classifier Model and Studying the Features and Applications of:
    a. Feature Engineering (Feature Normalisation, Feature Selection, Dimensionality Reduction, Best parameter for Classifiers using Grid Search applied on Stratified k-fold Cross Validation).
    b. Classifier – SVM, Ensemble Methods, Combination Rule (Majority Vote)
4. Highest Accuracy Reached – **99.22%**

❖ *Summary of the Techniques Used*:
1. Normalising the Features (*normalize*).
2. Dimensionality Reduction (*PCA*).
3. Feature Selection using LinearSVC (*SelectFromModel*).
4. Parameters for Classifier with Stratified 10-Fold Cross Validation (*StratifiedKFold*).
5. Classifiers Used:
    o Random Forest (*RandomForestClassifier*) – 97.94%
    o Adaboost (*AdaBoostClassifier*) – 67.78%
    o Gradient Boosting (*GradientBoostingClassifier*) – 86.5%
    o Majority Voting (*VotingClassifier*) – 98.72%

Model Presented: - Feature Engineering using Normalize, PCA, & SelectFromModel, Stratified 10-Fold Cross Validation and training a SVC and getting the confusion matrix an accuracy of *99.22%*.

Note: AdaBoost contrary to its theoretical innards gave performances <30% until sincere hyper parameter tuning. On the other hand, basic Feature Engineering with SVM gives significantly better results. Majority Voting impressively works well, and its fine-tuning gives an idea of the democracy amidst the classifiers.

Silly Note: The stealth of Grid Search could be eased with Parallel Computing (as per discussions).