Project 1 – Iris Dataset

Context: This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

Dataset Source: click here

Task: Predict the class of iris plant.

Questions to Answer:

- Perform EDA and Explore the features using histograms
- Encode the target variable ie convert it to numeric type
- Experiment using two different ratios of training, validation and test data ie 60-20-20, 80-10-10. On the two different split ratios do the following
    - Implement KFold Cross Validation
    - Implement Grid Search to find optimal hyperparameters for any 3 algorithms (out of LR, SVM, MLP, RF, Boosting)
    - Analyze the results on Validation set and test set and mention which model performed the best and why?
    - Compare the performance of models(using precision, recall, accuracy, latency). What was the best proportion or split ratio of data from the set of experiments you conducted?

Submission Instructions: Please just submit one jupyter notebook containing all the code and make use of markdown cells to include the comments, answers, reasoning, analysis, etc.

Note: Name of your file should be your "Project1-id_Firstname_Lastname.ipynb"