

PROBLEM STATEMENT- PHARMA

Maple Leaves Ltd is a start-up company which makes herbs from different types of plants and its leaves. Currently the system they use to classify the trees which they import in a batch is quite manual. A labourer from his experience decides the leaf type and subtype of plant family. They have asked us to automate this process and remove any manual intervention from this process.

Objective: To classify the plant leaves by various classifiers from different metrics of the leaves and to choose the best classifier for future reference.

Implementation:

- 1) Import the train and test csv.
- 2) Import the required classification libraries along with pandas, numpy, seaborn etc
- 3) Then import the classifiers from them (Randomforest, SVM, NaiveBayes, DecisionTrees)
- 4) After this create a function to encode the labels of the strings given in the dataset
- 5) You can do the above step using label encoder. With this you are creating some labels from train set as test set. The test set we imported is for testing the best classifier accuracy once we choose it
- 6) Then extract the values from train set by stratifying them and dividing it into 80:20 ratio
- 7) Now your X train, X test, Y train, Y test are ready.
- 8) We currently don't know which is the best classifier on the dataset. So, we apply all 4 of them.
- 9) Create the classifiers class and initialize all the respective classifiers
- 10) Then run the X train & X test datasets through classifiers calculating the log loss and accuracy of the result
- 11) Choose the classifier which has the best accuracy
- 12) Then try to predict the result on the import test.csv dataset