Project

# Platform:

Anaconda Notebook

# Language:

Python

Libraries:

NumPy, Pandas, Matplotlib, Scikit\_learn

These are the minimum requirements for the project. We may have to use other libraries in future.

Input Data for Training (mainly in csv or excel file)

Split the data (Training set and Test Set) ~70% and 30% division approximately (these values are subjected to change according to the need of data)

Test set data will be used for internal accuracy checking

As the input file contains categorical values (i.e. discrete in nature) we will be using classification which is as supervised technique.

Different Classification algorithms are there, like logistic regression, Nearest Neighbor, Support Vector Machine, Decision Trees, Boosted Trees, Random forest, Neural Networks etc.

With that classifier we will test accuracy for test data (from original input data) and modify the parameter values accordingly in previous steps.

Finally, we will try to predict the output for a given unknown data.