Development of an automated ML-based tool for classification problems

Aim: To develop a comprehensive python script that will run (binary of multi-class) classification problems on <u>any</u> given input data matrix

Input matrix Format: Standard Feature x Instance matrix (.csv) file.

	F1	F2	F3	F4	...	...	Fm	Class
I1								Class1
I2								Class2
I3								Class3
I4								
I5								
....								
In								Class x, y, z

The options for the following functions or subroutines will be asked from the user as standard input

Rules:

1. Every script(s) should contain elements of an ideal machine learning pipeline –

   a. <u>Normalization / standardization techniques</u> Atleast 3 options of normalization / standardization using different normalization techniques from which user will be given option to choose
   b. <u>Feature selection</u>: Atleast 5 options of selection techniques belonging to dimensionality reduction techniques and filter/wrapper techniques from which user can choose from.
   c. Cross validation: Atleast 3 options for cross validation techniques which can be taken as input from user
   d. <u>Machine learning models</u>. Atleast options for 5 different modelling (classification) techniques should be given to the user
   e. Predictive capability (accuracy) of the model should be validated using a <u>blind dataset</u> that can be 10% of the data which will be kept aside and will never be used for training or testing of the models.
   f. All performance metrics should be reported for overall cross-validated sets

2. <u>Output of the scripts</u> (plots, tables, printing) should be in the form of a **pdf file**: Think of printing and displaying output in an organized manner.