**Methodology:**

* Get the list of data files mentioned in the datasets\_selection.
* With this, read the config information for each data file.
* Save the information in separate lists, for ex, separatorList, headerList etc.,
* Convert **string lists** of category and value indices to **int list** (subtract 1 for index in python) by passing them through strSeparator function.
* Create an empty data frame ‘finalDf’ to store the final result with columns for data file name and the accuracy score for each algorithm.
* Loop through the data file list and for each file name, change the directory to the corresponding folder and read the data file as a csv using the config information.
* Handle the missing values ‘?’ using na\_values and drop rows containing ‘na’ values.
* Loop through the columns in the data file to rename them based on whether they are categorical (cat) or value (num) or target (out) columns.
* Drop columns that are not categorical/value/target.
* Pass this dataframe through pipeline transformation functions defined earlier.
* Function pre-processes the input variables by encoding categorical data in the one-hot dense format, scales the numerical variables and standardize all variables.
* It also imputes for the missing data by filling them with median values.
* Add the target column to this transformed dataframe. This is the dataframe on which the algorithms will be run.
* Split the dataframe into train (80%) and test (20%) data sets.
* Initialize the input and output variables.
* For each classifier, define the classifier and create a fit using the train data.
* Compute the accuracy of this fit on the test data and store it in a variable.
* Insert a new row containing the dataFile name and accuracy values into the Final table ‘finalDf’.
* Change directory back to the original directory.
* Go to the next iteration in the dataFile list.
* Write the finalDf table to a csv named ‘FinalDF’.

Few rows of FinalDF:

