**Classification Report**

**KNN:**

* Read the data from the file ‘census1994.csv’
* Print the first and last 5 rows using data.head() and data.tail() functions
* Remove attributes with ‘?’ using the ‘dropna()’ function
* Select only 2 attributes from the whole data. I have selected ‘education-num’ and ‘occupation’ as the attributes since these gave the highest accuracy
* Split the data into training and test with 70% as the train size
* Create KNN classifier using KNeighborsClassifier from sklearn with Euclidean as the distance metric and K values as [3,5,7]
* Create Confusion matrix and classification report from sklearn.metrics
* Plot the accuracies for all the models using matplotlib.pyplot

**Decision Tree:**

* Read the data from the file ‘census1994.csv’
* Print the first and last 5 rows using data.head() and data.tail() functions
* Remove attributes with ‘?’ using the ‘dropna()’ function
* Convert Nominal data to ordinal data using pandas.get\_dummies(data)
* Split the data into training and test with 70% as the train size
* Create DT classifier using DecisionTreeClassifier from sklearn with max\_depth=5, min\_impurity\_split =0.3 and GINI and ENTROPY as the criterion for information
* Create Confusion matrix and classification report from sklearn.metrics
* Create a tree\_to\_code() function the gets all the rules from the DecisionTreeClassifier
* Plot the decision tree using sklearn.tree.plot\_tree
* Plot the accuracies for all the models using matplotlib.pyplot

**References:**

<https://stackoverflow.com/questions/20224526/how-to-extract-the-decision-rules-from-scikit-learn-decision-tree>

<https://scikit-learn.org/>