**Classification Assignment (CKD dataset)**

Problem statement:

A requirement from the Hospital, Management asked us to create a predictive model which will predict the Chronic Kidney Disease (CKD) based on the several parameters. The Client has provided the dataset of the same.

Stage -1 : Machine learning

Stage -2: Supervised Learning

Stage-3: Classification.

1.) To predict the Chronic Kidney Disease (CKD)

2.) Tell basic info about the dataset (Total number of rows, columns) :

Rows: 399 and Columns: 25

3.) Mention the pre-processing method if you’re doing any (like converting

string to number – nominal data)

converted the nominal data for the columns with string values.

SC – Scalar method for pre-processing

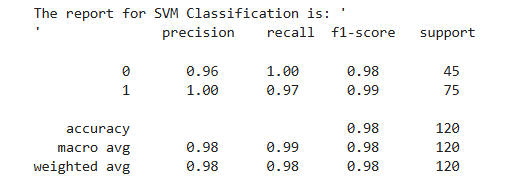
4.) Create a good evaluation metric, using machine learning algorithm; create many models. Finally, to come up with the Best suitable model.

5.) All the research rules of each algorithm should be documented. (You

can make tabulation or screenshot of the results.)

6.) Mention your final model, justify why u have chosen the same.

1. SVM Classification – Grid:

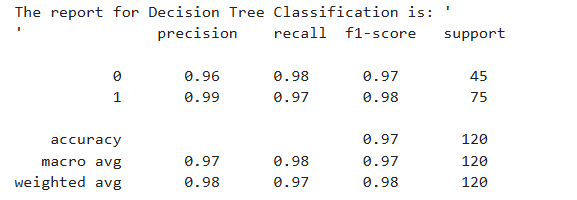


' [[45 0]

[ 2 73]]

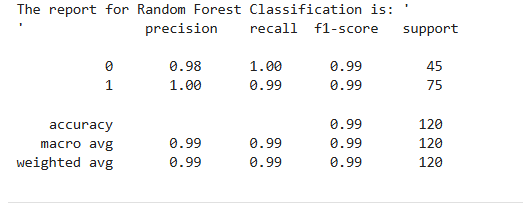
* 1. Accuracy : 98%
  2. Best parameters found: {'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid'
  3. F1Score : 0.98
  4. roc value : 0.99

1. Decision Tree Classification – Grid:



* 1. Accuracy : 97%
  2. Best parameters found: {'criterion': 'gini', 'max\_features': 1, 'splitter': 'best'}
  3. F1Score : 0.975
  4. roc value : 0.975

1. Random Forest Classification - Grid:

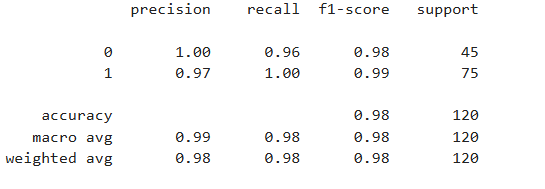


' [[45 0]

[ 1 74]]

* 1. Accuracy : 99%
  2. Best parameters found: {'criterion': 'entropy', 'max\_features': 1, 'n\_estimators': 100}
  3. F1Score : 0.99
  4. roc value : 0.99

1. Logistic Regression – Grid Search

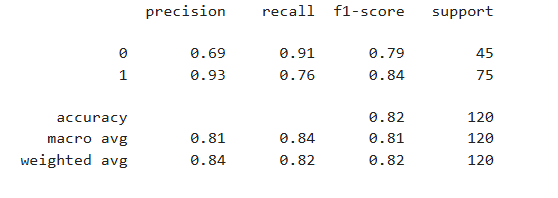


[[43 2]

[ 0 75]]

* 1. Accuracy : 0.98
  2. {'C': 100, 'penalty': 'l2', 'solver': 'liblinear'}
  3. F1Score : 0.98
  4. roc value : 0.99

1. KNN Classification

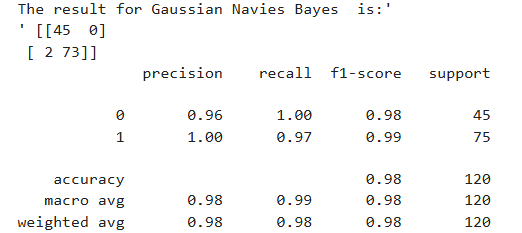


[[41 4]

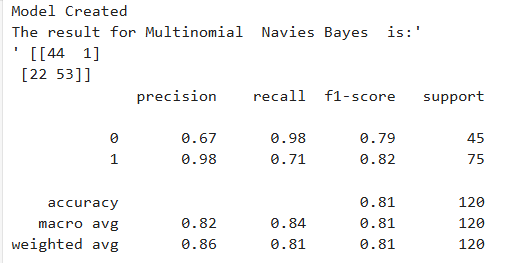
[18 57]]

* 1. Accuracy : 0.82
  2. Best parameters found: {'metric': 'manhattan', 'n\_neighbors': 1, 'weights': 'uniform'}
  3. Best cross-validation score: 0.7707792207792209
  4. F1Score : 0.819
  5. roc value : 0.83

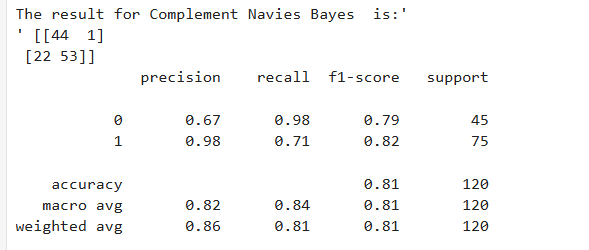
1. **Navies Bays Classification**



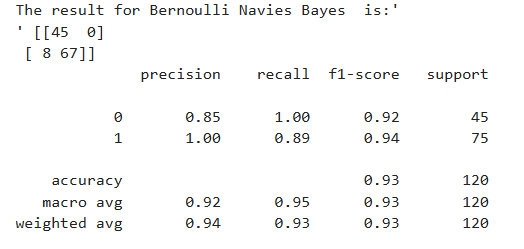
* 1. Accuracy : 98%
  2. Best parameters found: NA
  3. Best cross-validation score: NA
  4. F1Score : 0.98
  5. roc value : NA



1. Accuracy : 81%
2. Best parameters found: NA
3. Best cross-validation score:
4. F1Score : 0.82
5. roc value : NA



1. Accuracy : 81%
2. F1Score : 0.82
3. roc value :



1. Accuracy : 93%
2. F1Score : 0.94

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The best Model selected:

Random Forest Classification : Has the highest accuracy and F1 score . auc is almost 1

* 1. Accuracy : 99%
  2. Best parameters found: {'criterion': 'entropy', 'max\_features': 1, 'n\_estimators': 100}
  3. F1Score : 0.99
  4. roc value : 0.99