Hope Artificial Intelligence Classification Assignment

Problem Statement or Requirement:

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.

1.) Identify your problem statement

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

3.) Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)

4.) Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

5.) All the research values 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.

Answers:--

1.)Problem Statement: \* Develop a predict model to classify Chronic Kidney Disease (CKD) based on patient attributes.

2.)The dataset (Total number of rows, columns)

\* Total number of rows: 399

\* Total number of columns: 25

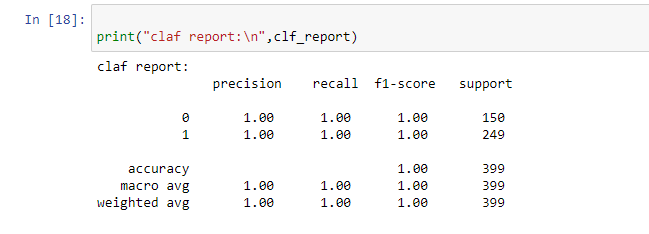
3.) The pre-processing method if you’re doing any (like converting s string to number – nominal data .

\*Convert categorical variables to numerical using one-hot encoding.

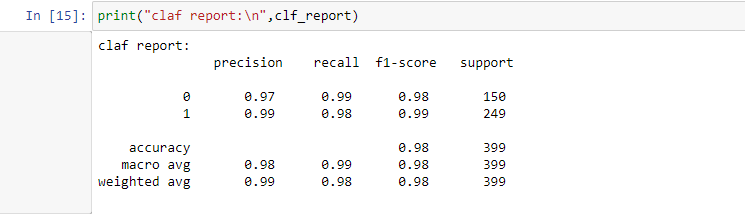
\*Convert the dataset to integer type.

4) All the research values of each algorithm should be documented. (You can make tabulation or screenshot of the results.

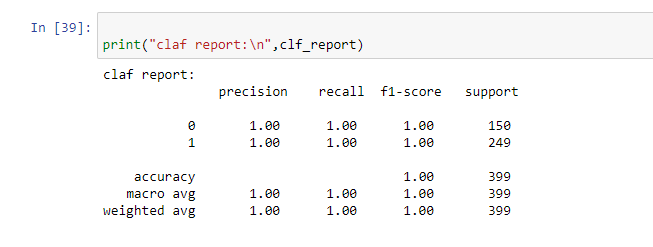
1. SVM:



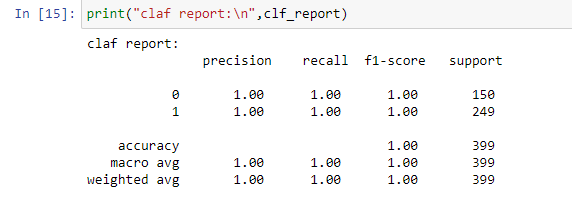
1. Decision tree:



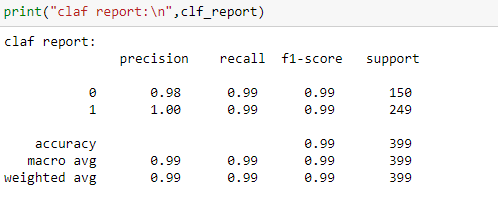
1. Random forest:



1. Logistic regression:-

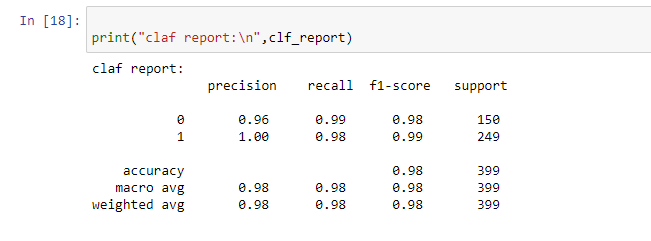


e)KNN:

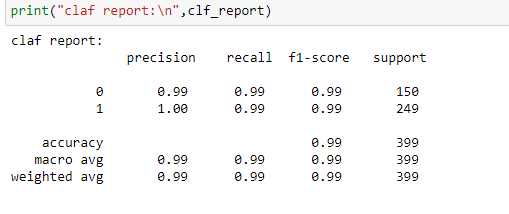


1. Navies’Bayes:

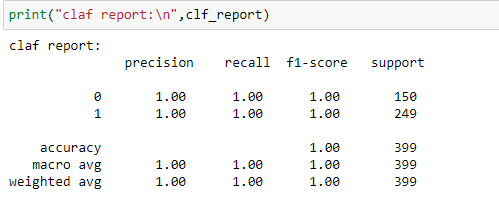
* GaussianNB:--



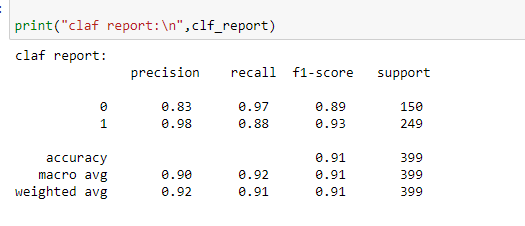
* BernoullisNB:--



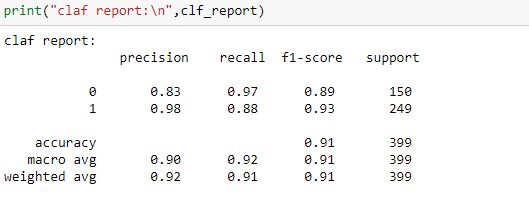
* CategoricalNB:



* MultinomialNB:--



* ComplimentNB:--



5) Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model .

Algorithm used: Support Vector Machine (SVM) , Random forest(RF), Logistic regression

Evaluation metric: Confusion matrix and classification report (including precision, recall, and F1-score) ,roc-auc score,Table report

Preprocessing: Standardization of input features using StandardScaler

Best Model parameters:

SVM== C=10, kernel='poly', degree=3, gamma='auto', coef0=0.0, random\_state=0

RF== Criterion: 'entropy', Max\_features: 'log2', n\_estimators: 100

Logistic regression==C: 1.0, Class\_weight: None, Multi\_class: 'auto', Penalty: 'l2'}

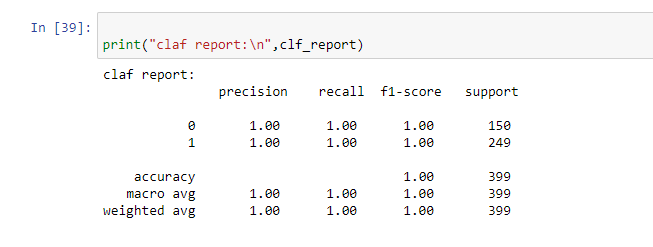
**Model Evaluation:**

Confusion matrix: Provides a tabular representation of actual & predicted classifications.

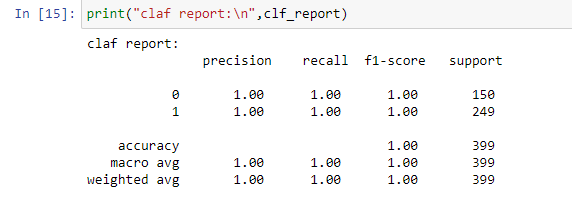
Classification report: Precision, recall, F1-score, and support for each class.

GOOD MODELS:

1]Random forest:-



2]Logistic regression:-



Here are the some algorithms gives 100% output as like SVM.So finally we selected SVM as a good model.

Class 0: Precision, Recall, and F1-score are 1.00, meaning correct predictions for this class.

Class 1: Precision ,Recall and F1-score is 1.00, indicating very good predictions for this class.

Accuracy of the model is also 1.00

Macro Avg: Average precision, recall, and F1-score across both classes.

Weighted Avg: Weighted average precision, recall, and F1-score considering the number of instances in each class

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

\*SVM was chosen because it can effectively handle high-dimensional data and create complex decision boundaries. The parameters for the SVM model were selected based on experimentation. Techniques like grid search or random search may have been used to adjust these parameters for better performance.

Final Model:

\* The final model is a Support Vector Machine classifier with the Following parameters:C=10,Kernel='poly',,gamma-‘auto’..Other parameters are default .