Assignment-Classification Algorithm

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.

As a data scientist, you must develop a model which will predict the insurance charges.

1. **Identify your problem statement**
   a. Stage 1 - Machine Learning
   b. Stage 2 - Supervised Learning
   c. Stage 3 – Classification
2. **Tell basic info about the dataset (Total number of rows, columns)**
   a. Total number of rows = 395
   b. Total number of columns = 25
   c. Input Columns = age, bp, al, su, bgr, bu, sc, sod, pot, hrmo, pcv, wc, rc, sg, rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane
   d. Output Column = classification
3. **Mention the pre-processing method if youre doing any (like converting string to number – nominal data)**
   e. As Input fields (rbc, pc, pcc, ba, htn, dm, cad, appet, pe and ane) and Output field (classification) are Categorical Nominal data, we need to pre-processing the dataset by converting those field values into number using One hot Encoding method.

4. **Classification Result based on Multiple algorithm**

   a. **Logistic Grid Classification**

The report :

	precision	recall	f1-score	support
0	0.98	1.00	0.99	51
1	1.00	0.99	0.99	82
accuracy			0.99	133
macro avg	0.99	0.99	0.99	133
weighted avg	0.99	0.99	0.99	133

Logistic Grid Classification roc_auc value = 1.0 and best f1_weighted value = 0.9924946382275899 ('penalty': 'l2', 'solver': 'newton-cg')

b. Support Vector Machine (SVM)

The report :

	precision	recall	f1-score	support
0	0.98	1.00	0.99	51
1	1.00	0.99	0.99	82
accuracy			0.99	133
macro avg	0.99	0.99	0.99	133
weighted avg	0.99	0.99	0.99	133

SVM Grid Classification roc_auc value = 0.9539015606242497 and best f1_weighted value = 0.9924946382275899 ('C': 10, 'gamma': 'auto', 'kernel': 'sigmoid')

c. Decision Tree

The report:

	precision	recall	f1-score	support
False	0.88	1.00	0.94	51
True	1.00	0.91	0.96	82
accuracy			0.95	133
macro avg	0.94	0.96	0.95	133
weighted avg	0.95	0.95	0.95	133

Decision Tree Classification roc_auc value = 0.9573170731707317 and the best f1_weighted value = 0.9478851104269762 ('criterion': 'entropy', 'max_features': 'log2', 'splitter': 'random')

d. **Random Forest**

The report:

	precision	recall	f1-score	support
False	0.94	1.00	0.97	51
True	1.00	0.96	0.98	82
accuracy			0.98	133
macro avg	0.97	0.98	0.98	133
weighted avg	0.98	0.98	0.98	133

Random Forest Regression roc_auc Value = 0.9573170731707317 and the best f1_weighted value = 0.9775556904684072 ('criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 10)

5. **The final machine learning best method of Classification:**
   Logistic Classification with **f1_weighted value ('penalty': 'l2', 'solver': 'newton-cg') =** **0. 9924946382275899** and **the roc_auc value = 1.0**