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.

1. Identify your problem statement

Name of the Project: Chronic Kidney Disease Prediction

Three stages of Problem Identification:

1.) Stage 1-**Machine Learning** (Numeric data with ordinal values)

2.) Stage 2-**Supervised** (input and output are clearly given)

3.) Stage 3-**Classification** (Categorical values with yes/no)

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

- This dataset contains 399 rows × 25 columns.
- It contains 27 independent (i/p) variables and 1 dependent (o/p) variable.

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

One-Hot Encoding has been done in pre-processing step to convert **categorical** data into a **numerical** format by using **get_dummies**.

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.

Created model using Classification Algorithms:-

- **Logistic Regression**
- **Random Forest Classifier**
- **Decision Tree Classifier**
- **K Nearest Neighbour**
- **Support Vector Machine Classifier**
- **Gaussian Naïve Bayes Classifier**
- **Complement Naïve Bayes Classifier**
- **Multinomial Naïve Bayes Classifier**

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

Logistic Regression: ROC_AUC_SCORE is 0.9997608799617408

The Report is:

	precision	recall	f1-score	support
False	0.96	1.00	0.98	51
True	1.00	0.98	0.99	82
accuracy			0.98	133
macro avg	0.98	0.99	0.98	133
weighted avg	0.99	0.98	0.99	133

Support Vector Machine Classifier: ROC_AUC_SCORE is 1.0

The Report is:

	precision	recall	f1-score	support
False	0.98	1.00	0.99	51
True	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

Decision Tree Classifier: ROC_AUC_SCORE is 0.9901960784313725

The Report is:

	precision	recall	f1-score	support
False	1.00	0.98	0.99	51
True	0.99	1.00	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

Random Forest Classsifier: ROC_AUC_SCORE is 0.9986848397895743

The Report is:

	precision	recall	f1-score	support
False	0.98	1.00	0.99	51
True	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

K Nearest Neighbour: ROC_AUC_SCORE is 0.9997608799617408

The Report is:

	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

Gaussian Naïve Bayes Classifier: ROC_AUC_SCORE is 1.0

The Report is:

	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

Multinomial NB: ROC_AUC_SCORE is 0.9151123864179818

The Report is:

	precision	recall	f1-score	support
False	0.68	0.98	0.81	51
True	0.98	0.72	0.83	82
accuracy			0.82	133
macro avg	0.83	0.85	0.82	133
weighted avg	0.87	0.82	0.82	133

Complement NB: ROC_AUC_SCORE is 0.9151123864179818

```
The Report is:
              precision    recall  f1-score   support

       False       0.68      0.98      0.81        51
        True       0.98      0.72      0.83        82

    accuracy                           0.82       133
   macro avg       0.83      0.85      0.82       133
weighted avg       0.87      0.82      0.82       133
```

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

MODEL	ROC_AUC_SCORE	ACCURACY
Support Vector Machine Classifier	**1.0**	**0.99**
Gaussian Naïve Bayes Classifier	**1.0**	**0.98**

- **Support Vector Machine Classifier** and **Gaussian Naïve Bayes Classifier** gives better **ROC_AUC_SCORE** than other models.
- Out of these two models, **Support Vector Machine Classifier** gives better accuracy than **Gaussian Naïve Bayes Classifier.**

Hence, the table above shows that **Support Vector Machine Classifier** is the most effective model to fit comparing to **Gaussian Naïve Bayes Classifier**.