Classification Assignment

1.) Identify your problem statement?
- Problem Statement was identified as Machine Learning.
- Under Machine Learning it was a Supervised Learning since we having Input and Output Data in the Data set
- In ML we are going to User classification since it was having a Categorical Datatype(Yes/No)
- Our Prediction/Output is going to check CKD is Yes/No based on the Given Inputs

2.) Tell basic info about the dataset (Total number of rows, columns)
- 399 rows × 25 columns

3.) Mention the pre-processing method if you're doing any (like converting string to number nominal data)
- Since it was classification Statement we need to convert the Categorical Data into Numerical 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.
- Random Forest Classification
- Decision Tree – Classification
- SVM- Classification
- Logistic Regression
- KNN
- NVB
- Logistic Grid Classification
- SVM Grid Classification
- Decision Tree Grid Classification

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

i. Random Forest Classification – 98%

	precision	recall	f1-score	support
False	0.96	0.98	0.97	51
True	0.99	0.98	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

ii. Decision Tree – Classification -92%

	precision	recall	f1-score	support
False	0.84	0.96	0.90	51
True	0.97	0.89	0.93	82
accuracy			0.92	133
macro avg	0.91	0.93	0.91	133
weighted avg	0.92	0.92	0.92	133

iii. SVM- Classification – 62%

	precision	recall	f1-score	support
False	0.00	0.00	0.00	51
True	0.62	1.00	0.76	82
accuracy			0.62	133
macro avg	0.31	0.50	0.38	133
weighted avg	0.38	0.62	0.47	133

iv. Logistic Regression – 89%

	precision	recall	f1-score	support
False	0.82	0.90	0.86	51
True	0.94	0.88	0.91	82
accuracy			0.89	133
macro avg	0.88	0.89	0.88	133
weighted avg	0.89	0.89	0.89	133

v. KNN-69%

	precision	recall	f1-score	support
False	0.57	0.78	0.66	51
True	0.83	0.63	0.72	82
accuracy			0.69	133
macro avg	0.70	0.71	0.69	133
weighted avg	0.73	0.69	0.70	133

vi. NB-MultinomialNB -79%

	precision	recall	f1-score	support
False	0.65	0.98	0.78	51
True	0.98	0.67	0.80	82
accuracy			0.79	133
macro avg	0.82	0.83	0.79	133
weighted avg	0.85	0.79	0.79	133

NB-

vii. Logistic Grid Classification – 98%

The report:

	precision	recall	f1-score	support
False	0.96	0.98	0.97	51
True	0.99	0.98	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

viii. SVM-Grid Classification – 91%

The report:

	precision	recall	f1-score	support
False	0.85	0.92	0.89	51
True	0.95	0.90	0.93	82
accuracy			0.91	133
macro avg	0.90	0.91	0.91	133
weighted avg	0.91	0.91	0.91	133

ix. DecisionTree.Grid Classification – 93%

The report:

	precision	recall	f1-score	support
False	0.92	0.90	0.91	51
True	0.94	0.95	0.95	82
accuracy			0.93	133
macro avg	0.93	0.93	0.93	133
weighted avg	0.93	0.93	0.93	133

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

- Random Forest Classification is the Best model, since it was having best Accuracy of 98% and also compared with the data set it was providing the correct result in the Deployment phase.