Classification Assignment

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

To predict the Chronic Kidney Disease based on the parameters given in the dataset

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

There are 399 rows and 28 columns . They are 'age', 'bp', 'al', 'su', 'bgr', 'bu', 'sc', 'sod', 'pot', 'hrmo', 'pcv', 'wc', 'rc', 'sg_b', 'sg_c', 'sg_d', 'sg_e', 'rbc_normal', 'pc_normal', 'pcc_present', 'ba_present', 'htn_yes', 'dm_yes', 'cad_yes', 'appet_yes', 'pe_yes', 'ane_yes', 'classification_yes'.

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

Using get_dummies the string data of the column is converted into 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.
The final model roc_auc_score

Logistic:1.0
SVM:1.0

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

Logistic regression:

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

RF grid

The report:

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

Decision Tree Grid

The report:

	precision	recall	f1-score	support
0	0.89	0.96	0.92	51
1	0.97	0.93	0.95	82
accuracy			0.94	133
macro avg	0.93	0.94	0.94	133
weighted avg	0.94	0.94	0.94	133

SVM Grid

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

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

The model using Logistic regression algorithm is my final model. Though the roc_auc_score of SVM and Logistic regression is same , Logistic is chosen as final model because the model creation in logistic is easier than SVM