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:***

   Medical Analytics for CKD Prediction

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

   The dataset CKD.csv contains the following basic information:

   - Total number of rows: **399**
   - Total number of columns: **25**

   The dataset includes various medical attributes such as age, blood pressure (bp), specific gravity (sg), albumin (al), sugar (su), and a classification column among others.

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

Pre-processing Summary Table		
Feature Type	**Original Data (Example)**	**Pre-processed Data**
Nominal (Binary)	"normal", "abnormal"	1, 0
Nominal (Multi-class)	"a", "b", "c", "d", "e"	Multi-column binary flags
Target Variable	"yes" (CKD), "no" (Not CKD)	1, 0
Numeric	76.45, 148.11	Kept as Float/Int

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***

Algorithm	Best Use Case	Complexity	Status
SVM (Grid)	High-dimensional data (excellent for TF-IDF).	High	Top Performer
Naive Bayes	Large datasets with many word features; very fast.	Low	Highly Efficient
Random Forest (Grid)	Reducing variance; handles noisy text well.	High	Strong & Robust
Logistic Regression	Linear relationships; fast baseline.	Low	Reliable Baseline
KNN (Grid)	Finding "neighboring" similar document styles.	Medium	Fair
Decision Tree (Grid)	Highly interpretable "if-then" logic.	Medium	Interpretability

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

- ***Logistic Regression Grid Model:***

	precision	recall	f1-score	support
0	0.94	0.98	0.96	51
1	0.99	0.96	0.98	82
accuracy			0.97	133
macro avg	0.97	0.97	0.97	133
weighted avg	0.97	0.97	0.97	133

- ***Decision Tree Grid Model:***

	precision	recall	f1-score	support
0	0.96	0.94	0.95	51
1	0.96	0.98	0.97	82
accuracy			0.96	133
macro avg	0.96	0.96	0.96	133
weighted avg	0.96	0.96	0.96	133

- ***Random Forest Grid Model:***

	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

- ***Support Vector Machine Grid Classification:***

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

- ***KNN Grid Model:***

	precision	recall	f1-score	support
0	0.69	0.94	0.79	51
1	0.95	0.73	0.83	82
accuracy			0.81	133
macro avg	0.82	0.84	0.81	133
weighted avg	0.85	0.81	0.81	133

- ***Naïve Bayes Grid Model:***

	precision	recall	f1-score	support
0	0.94	1.00	0.97	51
1	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

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

Rank	Model	Macro F1-Score	Status
1	Support Vector Machine (SVM)	1	Final Model
2	Naive Bayes	0.98	Excellent
3	Random Forest	0.98	Excellent
4	Logistic Regression	0.97	Very Good
5	Decision Tree	0.86	Good
6	KNN (K-Nearest Neighbors)	0.82	Fair

****