1. Problem Statement

To create a model to predict Chronic Kidney Disease.

2. Basic Info About the Dataset

The dataset has 399 rows and 25 columns. It has the complete details about a patient to predict the Chronic Kidney Disease occurrence in future.

3. Pre-Processing Method

The dataset has string data, so to convert the nominal data to ordinal data, **one-hot encoding** has been used in the dataset.

4. Algorithms used:

- Decision Tree
- Random Forest Classifier
- SVC
- Logistic Regression
- KNNeighbors
- Naive Bayes
- GaussianNB
 - MultinomialNB
 - ComplementNB
 - o BernoulliNB

5. Algorithm values:

Algorithm Name	confusion_ma trix		classifica	Best Parameter			
Decision Tree	[[44 1] [0 75]]	0 1 accuracy macro avg weighted avg	precision 1.00 0.99 0.99 0.99	necall 0.98 1.00 0.99 0.99	f1-score 0.99 0.99 0.99 0.99 0.99	support 45 75 120 120 120	{'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 200}

							1
			precision	recall	f1-score	support	
Random Forest							
		0 1	0.98 0.99	0.98 0.99	0.98 0.99	45 75	{'criterion': 'gini',
	[[44 1]		****	0,,,,	****	, ,	'max_features': 'log2',
Classifier	[1 74]]	accuracy			0.98	120	'n estimators': 500}
		macro avg weighted avg	0.98 0.98	0.98 0.98	0.98 0.98	120 120	n_estimators : 500}
		gcu uvg	0.20	0120		-120	
			precision	recall	f1-score	support	
			processen		,1 500,0	очрро, с	
		0	0.92 0.99	0.98 0.95	0.95 0.97	45 75	
svc	[[44 1]	1	0.99	0.95	0.97	/5	kernel= 'linear', gamma
	[4 71]]	accuracy			0.96	120	='scale',C = 10
		macro avg	0.95 0.96	0.96 0.96	0.96 0.96	120 120	
		weighted avg	0.96	0.96	0.96	120	
			precision	recall	f1-score	support	
		0	1.00	0.96	0.98	45	
Logiotio	[[40,0]	1	0.97	1.00	0.98	45 75	('nanalty': 'IO', 'lackyor':
Logistic	[[43 2]						{'penalty': 'l2', 'solver':
Regression	[0 75]]	accuracy	0.99	0.98	0.98 0.98	120 120	'liblinear'}
		macro avg weighted avg	0.98	0.98	0.98	120	
	[[41 4]		precision	recall	f1-score	support	
		0	0.61	0.91	0.73	45	
		1	0.92	0.65	0.73	75	n_neighbors=5,metric=
KNNeighbors	[26 49]]						'minkowski', p=2
	[20 10]]	accuracy	0.77	0.70	0.75	120	Timikowski, p 2
		macro avg weighted avg	0.77 0.81	0.78 0.75	0.75 0.75	120 120	
	1	T	Naive Bay	es			1
			precision	recall	f1-score	support	
		0	0.96	1.00	0.98	45	
GaussianNB	[[45 0]	1	1.00	0.97	0.99	75	
CaussiallivD	[2 73]]	accuracy			0.98	120	
		macro avg	0.98	0.99	0.98	120	
		weighted avg	0.98	0.98	0.98	120	
			precision	recall	f1-score	support	
						1.15	
		0	0.67	0.98	0.79	45	
MultinomialNB	[[44 1]	1	0.98	0.71	0.82	75	
indition and	[22 53]]	accuracy			0.81	120	
		macro avg	0.82	0.84	0.81	120	
		weighted avg	0.86	0.81	0.81	120	
							<u> </u>
	-	-					

			precision	recall	f1-score	support
		0	0.67	0.98	0.79	45
	[[44 1]	1	0.98	0.71	0.82	75
ComplementNB	[22 53]]					400
	الدح عماا	accuracy macro avg	0.82	0.84	0.81 0.81	120 120
		weighted avg	0.86	0.81	0.81	120
			precision	recall	f1-score	support
		0	0.85	1.00	0.92	45
	[[45 0]	1	1.00	0.89	0.94	75
BernoulliNB	[8 67]]	266117261			0.93	120
	[0 07]]	accuracy macro avg	0.92	0.95	0.93	120
		weighted avg	0.94	0.93	0.93	120
			precision	recall	f1-score	support
	F 4 5 0 3	0	0.98	1.00	0.99	45
	[[45 0]	1	1.00	0.99	0.99	75
CategoricalNB	[1 74]]	accuracy			0.99	120
		macro avg	0.99	0.99	0.99	120
		weighted avg	0.99	0.99	0.99	120

6. Final Model & Values

Algorithm	Decision Tree								
Confusion Matrix	[[44 1] [0 75]]								
Classification Report			precision	recall	f1-score	support			
		0	1.00	0.98	0.99	45			
		1	0.99	1.00	0.99	75			
	accur	асу			0.99	120			
	macro	avg	0.99	0.99	0.99	120			
	weighted	avg	0.99	0.99	0.99	120			
Best Parameter	{'criterion': 'gini', 'max_features': 'log2', 'n_estimators': 200}								

Reason for choosing Decision Tree as Best Algorithm for this dataset:

Accuracy is 99% and with 100% recall (no false negatives), high precision, and an excellent F1-score.