Assignment-Classification Algorithm

Problem Statement or Requirement:

As a data scientist, you must develop a model which will predict the CKD.

1 Identify your problem statement

1.Machine learning(input given number in datasheet)
2.supervised learning(input and output clear)
3.Classification(output given classifier)

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

1.Total number of rows-399
2.Total columns-25

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

dataset=pd.get_dummies(dataset,drop_first=True)

4 Develop a good model with confusion matrix. You can use any machine learning algorithm; you can

Algorithm: Decision Tree&, Random Forest and Logistic Regression
Confusion Matrix=0.99

5 All the research values (confusion matrix of the models) should be documented.

Comparative Performance Analysis of Classifier Algorithms

Algorithms:

Support Vector Machine					
Kernel	**C**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Linear	10	0.98	1	0.97	0.99
Linear	100	0.98	1	0.97	0.99
Linear	1000	0.98	1	0.97	0.99
Linear	2000	0.98	1	0.97	0.99
Linear	3000	0.98	1	0.97	0.99
Rbf	10	0.98	1	0.97	0.99
Rbf	100	0.98	1	0.97	0.99
Rbf	1000	0.98	1	0.97	0.99
Rbf	2000	0.98	1	0.97	0.99
Rbf	3000	0.98	1	0.97	0.99
Poly	10	1	1	1	1
Poly	100	0.98	1	0.97	0.99
Poly	1000	0.97	1	0.96	0.98
Poly	2000	0.97	1	0.96	0.98
Poly	3000	0.97	1	0.96	0.98
Sigmoid	10	0.98	1	0.97	0.99
Sigmoid	100	0.98	0.99	0.97	0.98

Sigmoid	1000	0.97	0.99	0.97	0.98
Sigmoid	2000	0.97	0.99	0.97	0.98
Sigmoid	3000	0.97	0.99	0.97	0.98
Grid_Search-Support Vector Machine					
Sigmoid	10	0.98	1	0.97	0.99

Decision Tree Classifier						
Criterion	**Splitter**	**max_features**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Gini	best	auto	0.93	0.97	0.91	0.94
Gini	best	sqrt	0.97	0.97	0.97	0.97
Gini	best	log2	0.94	0.96	0.95	0.95
Gini	random	auto	0.98	0.99	0.99	0.99
Gini	random	sqrt	0.96	1	0.93	0.97
Gini	random	log2	0.94	0.99	0.92	0.95
Entropy	best	auto	0.93	0.97	0.91	0.94
Entropy	best	sqrt	0.97	1	0.96	0.98
Entropy	best	log2	0.97	0.96	0.99	0.97
Entropy	random	auto	0.97	0.97	0.97	0.97
Entropy	random	sqrt	0.97	1	0.95	0.97
Entropy	random	log2	0.97	1	0.96	0.98
Grid_Search-Decision Tree Classifier						
Entropy	random	auto	0.97	0.97	0.97	0.97

Random Forest Classifier						
Criterion	**n_ esimators**	**max_features**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Gini	10	sqrt	0.98	0.99	0.99	0.99
Gini	100	sqrt	0.98	0.99	0.99	0.99
Gini	10	log2	0.98	0.99	0.99	0.99
Gini	100	log2	0.99	0.99	1	0.99
Entropy	10	sqrt	0.97	0.99	0.96	0.97
Entropy	100	sqrt	0.98	0.99	0.99	0.99
Entropy	10	log2	0.99	0.99	1	0.99
Entropy	100	log2	0.99	1	0.99	0.99
Grid_Search-Random Forest Classifier						
Entropy	100	sqrt	0.98	0.99	0.99	0.99

Logistic Regression					
penalty	**Solver**	**Accuracy**	**Precision**	**Recall**	**F1-score**
l2	Newton-cg	0.99	1	0.99	0.99
Grid_Search-Logistic Regression					
l2	Newton-cg	0.99	1	0.99	0.99

K-Nearest Neighbor						
n_ neighbors	**algorithm**	**metric**	**Accuracy**	**Precision**	**Recall**	**F1-score**
5	auto	minkowski	0.95	1	0.92	0.96
5	ball-tree	minkowski	0.95	1	0.92	0.96
5	kd-tree	minkowski	0.95	1	0.92	0.96
5	brute	minkowski	0.95	1	0.92	0.96
Grid_Search-K-Nearest Neighbor						
7	kd-tree	minkowski	0.95	1	0.92	0.96

Naive Bayes				
NB-models	**Accuracy**	**Precision**	**Recall**	**F1-score**
gaussianNB	0.98	1	0.97	0.99
multinominalNB	0.81	0.98	0.71	0.82
bernoulliNB	0.93	1	0.89	0.94
complementNB	0.81	0.98	0.71	0.82