Assignment-Classification Algorithms

1.Problem Statement –

Client (Hospital Management) wants us to create a model which predict chronic kidney disease (CKD) based on the several parameters given in the data set.

Domain Selection-Machine Learning.

Learning Selection-Supervised Learning-Classification

2.Basic info about the dataset-

Parameters-AGE, BP, SG, AL, SU, RBC, PC. PCC, BA, BGR, BU, SC, SOD, POT, HMO, PCV, WC, PCV, WC, RC, HTN, DM, CAD, APPET, PE, ANE, CLASSIFICATION.

Independent (Input)- AGE, BP, SG, AL, SU, RBC, PC. PCC, BA, BGR, BU, SC, SOD, POT, HMO, PCV, WC, PCV, WC, RC, HTN, DM, CAD, APPET, PE, ANE.

Dependent (Output)- CLASSIFICATION

Rows-400

Columns-25

3.Pre-Processing Method-

Parameter has staring values we are converting nominal data to number data.

Algorithm used-One hot Encoding

4.Model used -

Machine learning algorithm used a chronic kidney disease prediction.

Algorithms used-Random Forest, Support Vector Machine, Decision tree, Logistic, K-Nearest Neighbour, Naive Bayes

5.Documentation on Confusion Matrix and clf-report-

Random Forest-

```
43]: print(cm)

[[44  1]
 [ 1 74]]

44]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.98	0.98	0.98	45
1	0.99	0.99	0.99	75
accuracy			0.98	120
macro avg	0.98	0.98	0.98	120
weighted avg	0.98	0.98	0.98	120

Decision Tree-

```
print(clf_report)
```

	precision	recall	f1-score	support
0	0.84	0.96	0.90	45
1	0.97	0.89	0.93	75
accuracy			0.92	120
macro avg	0.91	0.92	0.91	120
weighted avg	0.92	0.92	0.92	120

```
print(cm)

[[43  2]
 [ 8 67]]
```

SVM-

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        45
           1       0.62      1.00      0.77        75

    accuracy                           0.62       120
   macro avg       0.31      0.50      0.38       120
weighted avg       0.39      0.62      0.48       120
```

```
print(cm)
```

```
[[ 0 45]
 [ 0 75]]
```

Logistic-

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        45
           1       1.00      0.99      0.99        75

    accuracy                           0.99       120
   macro avg       0.99      0.99      0.99       120
weighted avg       0.99      0.99      0.99       120
```

```
print(cm)
```

```
[[45  0]
 [ 1 74]]
```

KNN-

```
print(cm)
```

```
[[37  8]
 [27 48]]
```

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.58      0.82      0.68        45
           1       0.86      0.64      0.73        75

    accuracy                           0.71       120
   macro avg       0.72      0.73      0.71       120
weighted avg       0.75      0.71      0.71       120
```

Navies Bayes-

```
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        45
           1       1.00      0.97      0.99        75

    accuracy                           0.98       120
   macro avg       0.98      0.99      0.98       120
weighted avg       0.98      0.98      0.98       120

[[45  0]
 [ 2 73]]
```

5.Final Model-

Logistic Algorithm-Accuracy-0.99

Among trying all the algorithms Logistic Algorithm predicted the Overall performance –0.99 which good for the dataset given.

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        45
           1       1.00      0.99      0.99        75

    accuracy                           0.99       120
   macro avg       0.99      0.99      0.99       120
weighted avg       0.99      0.99      0.99       120
```

```
print(cm)
```

```
[[45  0]
 [ 1 74]]
```