Classification Assignment

Problem Statement

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.

Recognition Based On The Problem Statement

- From the given problem statement it identified that to predict the Chronic Kidney Disease from the given problem statement
- For the provided Data set, the Machine Learning Classification can be used to predict the Chronic Kidney Disease.
- Dataset contains 399 rows and 25 columns
- Dataset contains nominal data so it is pre-processed

Machine Learning Models

Decision Tree

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.91      0.91      0.91        85
           1       0.84      0.84      0.84        49

    accuracy                           0.88       134
   macro avg       0.87      0.87      0.87       134
weighted avg       0.88      0.88      0.88       134

              precision    recall  f1-score   support
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,grid.predict_proba(x_test)[:,1])
```

```
0.8713085234093638
```

Support Vector Classification

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        51
           1       0.62      1.00      0.76        82

    accuracy                           0.62       133
   macro avg       0.31      0.50      0.38       133
weighted avg       0.38      0.62      0.47       133
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,classifier.predict_proba(x_test)[:,1])
```

```
0.2551410808225729
```

Random Forest

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       1.00      0.98      0.99        51
           1       0.99      1.00      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
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,classifier.predict_proba(x_test)[:,1])
```

```
0.9997608799617408
```

Logistic Regression

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.90      0.90      0.90        51
           1       0.94      0.94      0.94        82

    accuracy                           0.92       133
   macro avg       0.92      0.92      0.92       133
weighted avg       0.92      0.92      0.92       133
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,classifier.predict_proba(x_test)[:,1])
```

```
0.9820659971305594
```

Navie's Baye's

```
print(clf_report)
```

```
              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
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,classifier.predict_proba(x_test)[:,1])
```

```
1.0
```

K-nearest Neighbor

```
print(clf_report)
```

```
              precision    recall  f1-score   support

           0       0.61      0.86      0.72        51
           1       0.89      0.66      0.76        82

    accuracy                           0.74       133
   macro avg       0.75      0.76      0.74       133
weighted avg       0.78      0.74      0.74       133
```

```
from sklearn.metrics import roc_auc_score
roc_auc_score(y_test,classifier.predict_proba(x_test)[:,1])
```

```
0.7967479674796748
```

CONCLUSION

Model	roc_auc_score	Over all accuracy
Support vector mechanics	0.2551410808225729	0.62
Decision Tree	0.929220468675275	0.92
Random Forest	0.9997608799617408	0.99
Logistic Regression	0.9820659971305594	0.92
K-nearest neighbor	0.7967479674796748	0.74
Navie's Baye's	1.0	0.98

Result

Random forest machine model has a good level of accuracy and roc_auc_score for the provided dataset.