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

**Machine learning - Domain**

**Supervised learning – Learning method**

**Classification - Algorithm**

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

**(399, 28)**

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

Categorical data into numerical data(one hot encoding)

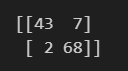
1. Develop a good model with good evaluation metric. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

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

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

Decision Tree:

confusion\_matrix:



classification\_report:

A screenshot of a computer

Description automatically generated

F1\_score:

0.93793

Roc\_auc\_score:

0.93111

Random forest algorithm:

confusion\_matrix:

A black background with white numbers

Description automatically generated

classification\_report:

A screenshot of a computer screen

Description automatically generated

F1\_score:

**0.986666**

Roc\_auc\_score:

**0.99970**

**Support vector Machine**

confusion\_matrix:

A black background with white numbers

Description automatically generated

classification\_report:

A screenshot of a computer

Description automatically generated

F1\_score:

**0.98666**

Roc\_auc\_score:

**0.9997**

**GaussianNB**

confusion\_matrix:

A black background with white text

Description automatically generated

classification\_report:

A screenshot of a computer screen

Description automatically generated

F1\_score:

**0.9864**

Roc\_auc\_score:

**1.0**

**Logistic Regression**

confusion\_matrix:

A black background with white text

Description automatically generated

classification\_report:

A screenshot of a computer screen

Description automatically generated

F1\_score:

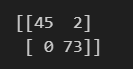
0.9868

Roc\_auc\_score:

0.9988

KNN

confusion\_matrix:



classification\_report:

A screenshot of a computer screen

Description automatically generated

F1\_score:

**0.98648**

Roc\_auc\_score:

**0.98666**

Choosing the best model

Random Forest is the best model and we got high accuracy for F1-score**0.986666**

, roc\_auc\_score-**0.999**