**Machine learning -Classification -GridCV-Assignment**

**Problem Statement or Requirement:**

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.) Problem Statement:

(**i) Machine Learning**

**(ii)Supervised Learning**

**(iii)Classification**

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

**399x25** (rows x columns) - before preprocessing

**399x28**(rows x columns) – after preprocessing

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

**One hot encoding method**

Random Forest Algorithm:

A screenshot of a computer

Description automatically generated

2.Decision Tree Algorithm:

A screenshot of a computer

Description automatically generated

3.SVM:

A screenshot of a computer

Description automatically generated

4.Logistic Regression:

A screenshot of a computer

Description automatically generated

5.KNN Algorithm:

A screenshot of a computer

Description automatically generated

6.(i)Naive Bayes:GaussianNB

A screenshot of a computer

Description automatically generated

6.(ii)Naives Bayes:Multinomial NB

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

6.(iii)Naives Bayes:Bernoulli NB:

A screenshot of a computer

Description automatically generated

6(iv)Naïve Bayes: Categorical NB:

A screenshot of a computer

Description automatically generated

6(v)Naïve Bayes: Complement NB:

A screenshot of a computer

Description automatically generated

Final Model:

Report of ML Classification algorithms using GridCV:

I had three best algorithms, RF, DT and KNN, I am choosing Random Forest as the best model.

The Best parameters for high accuracy was

**{'criterion': 'gini', 'max\_features': 'sqrt', 'n\_estimators': 100}**