**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. Identify your problem statement**

We need to predict the Chronic Kidney Disease based on the given parameters in the dataset. Machine Learning > Supervised Learning > Classification

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

Total number of columns:25

Total number of rows: 249

Independent/Input: 24 col

Dependent/Output: 1 col

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

used pd.get\_dummies which converted strings into int and standard scaler to optimize the difference between data.

**4. 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**.

Created classification models

1.SVM

2.Decision Tree

3. Logistic Regression

4. Random Forest

5.KNN

6. Naïve Bayes - Multinomial Naïve Bayes ,Categorical Naïve Bayes ,Complement Naïve Bayes ,Gaussian Naïve Bayes, Bernoulli Naïve Bayes.

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

The result screen shot below for ref,



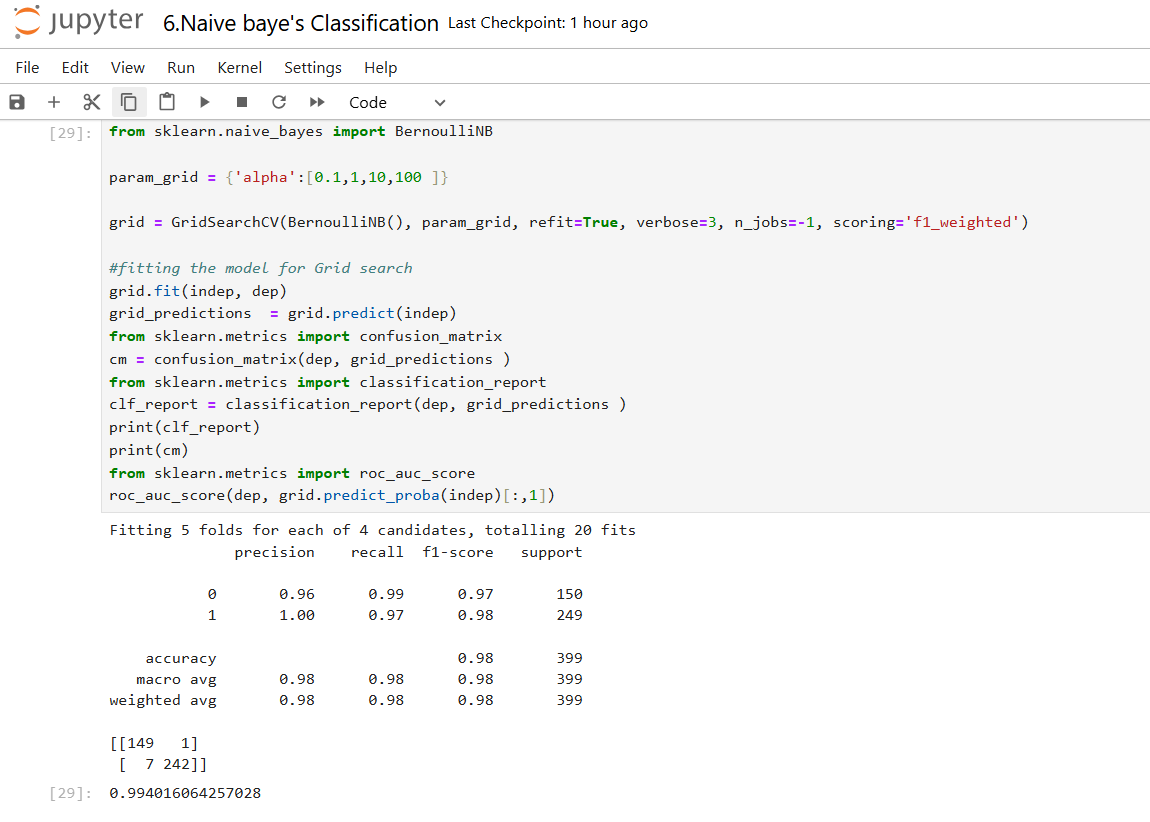


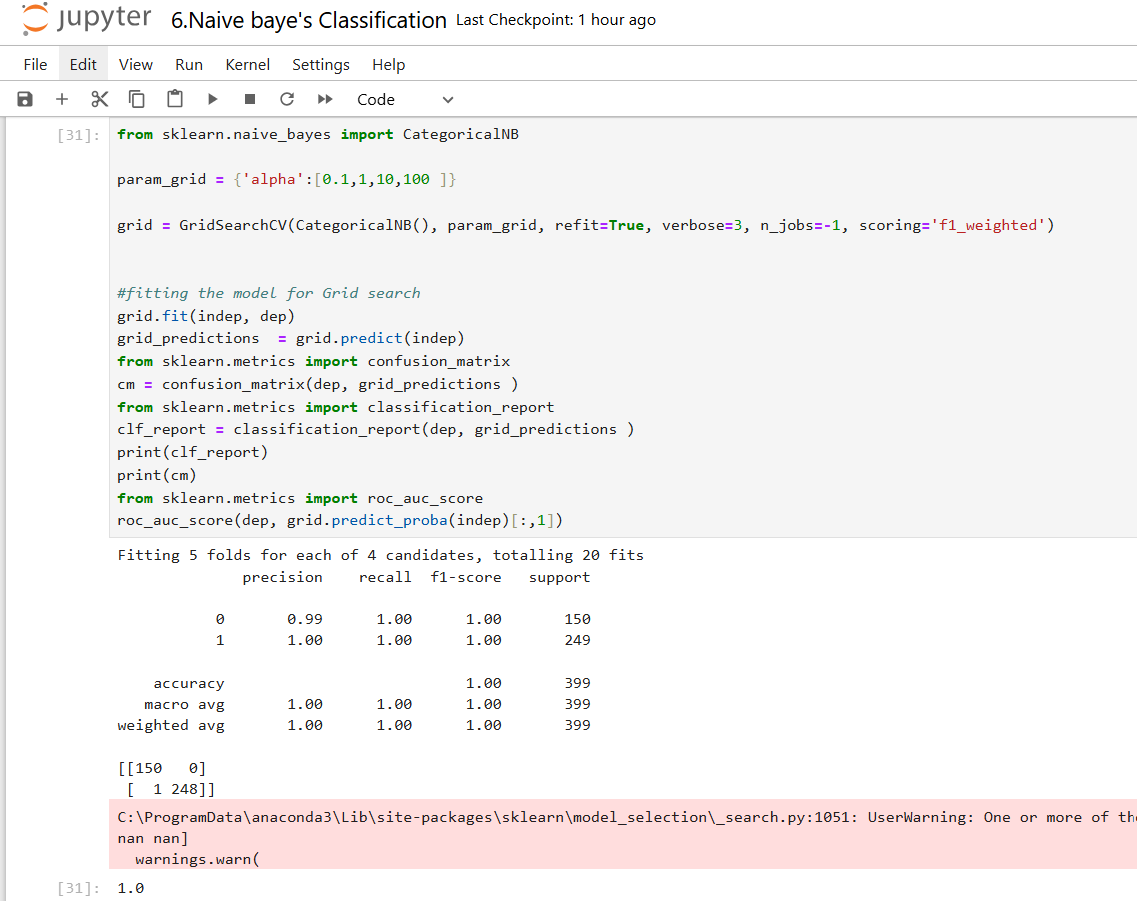


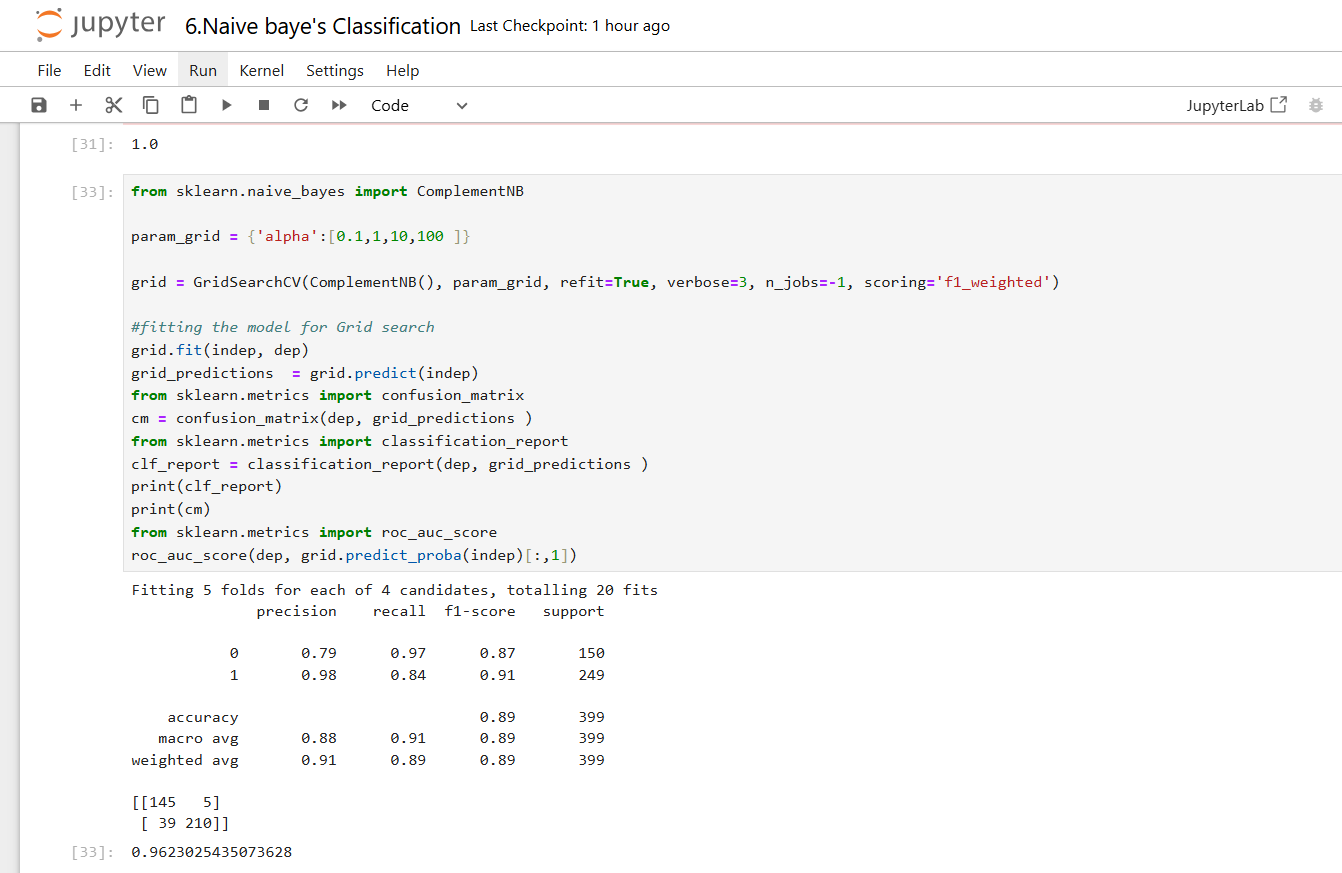














**6. Mention your final model, justify why u have chosen the same.**

Finally we used to select **Random Forest is best model** with below settings

The f1\_macro value for best parameter {'criterion': 'entropy', 'max\_features': 'log2', 'n\_estimators': 100}: 1.0