***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. The Client has provided the dataset of the same.

1. **Identify your problem statement**

This Comes Under **Machine Learning ,Classification and Supervised Learning**

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

Dataset:CKD.csv

Total No.of Rows : 399

Total No.of Columns : 25

After used get\_dummies-🡪 which change the categorical to numerical value

Total No.of Rows:399

Total No.of Columns : 28

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

Used **get\_dummies** to change the string to number (Nominal 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.**

* + - **RandomForest**
    - **Best Parameters(Using GridSearch)**
    - **Criterion:entropy,n\_estimators:500**

**5.) All the research values of each algorithm should be documented. (You**

**can make tabulation or screenshot of the results.)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S.no** | **Algorithm** | **Best Parameters**  **(Using GridSearch)** | **Precision** | **Recall** | **F1-score** | **Support** | **Accuracy** | **Roc**  **Auc**  **scor** |
| **1** | **Random Forest** | **Criterion:entropy,**  **n\_estimators:500** | **0-> 0.98**  **1-> 0.99** | **0.98**  **0.99** | **0.98**  **0.99** | **51**  **82** | **0.98** | **0.99** |
| **2** | **Decision**  **Tree** | Criterion:log\_  loss,splitter:  random | **0-> 0.98**  **1-> 0.99** | **0.98**  **0.99** | **0.98**  **0.99** | **51**  **82** | **0.98** | **0.98** |
| **3** | **SVM** | **C:3000,**  **kernel:linear** | **0->0.00**  **1->0.62** | **0.00**  **1.00** | **0.00**  **0.76** | **51**  **82** | **0.62** | **0.5** |
| **4** | **Logistic**  **Regression** | **C:1000,Penalty:l2,Slover:Saga** | **0->0.94**  **1->1.00** | **1.00**  **0.96** | **0.97**  **0.98** | **51**  **82** | **0.98** | **0.98** |
| **5** | **K\_Nearest\_Negibour** | **Metrics:manhattn,n\_neighbors:11,weights:uniform** | **0->0.88**  **1->1.00** | **1.00**  **0.91** | **0.94**  **0.96** | **51**  **82** | **0.95** | **0.95** |
| **6** | **Navies\_**  **Bayes** |  |  |  |  |  |  |  |
| **i** | **GaussianNB** | {'priors': None, 'var\_smoothing': 1e-05} | **0->0.94**  **1->1.00** | **1.00**  **0.96** | **0.97**  **0.98** | **51**  **82** | **0.98** | **0.98** |
| **ii** | **MultinomialNB** | **Alpha:1.0,priors=none,fit\_prior:True** | **0->0.69**  **1->0.98** | **0.98**  **0.73** | **0.81**  **0.84** | **51**  **82** | **0.83** | **0.85** |
| **iii** | **BernoulliNB** | **Alpha:1.0,binarize0.0,prior:True** | **0->0.85**  **1->1.00** | **1.00**  **0.89** | **0.92**  **0.94** | **51**  **82** | **0.93** | **094** |
| **iv** | **CategoricalNB** | **Not suitable** | **-**  **1->** | **-** | **-** | **-** | **-** | **-** |
| **v** | **ComplementNB** | **Alpha:1.0,priors=none,fit\_prior:True** | **0->0.69**  **1->0.98** | **0.98**  **0.73** | **0.81**  **0.84** | **51**  **82** | **0.83** | **0.85** |

***Final Model :*RandomForest**