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

Stage 1 -Domain selection- Machine Learning

Stage 2-Supervised

Stage 3-Classification

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

(399, 25)

* **Age(numerical):** age in years
* **Blood Pressure(numerical):** bp in mm/Hg
* **Specific Gravity(nominal):** sg - (1.005,1.010,1.015,1.020,1.025)
* **Albumin(nominal):** al - (0,1,2,3,4,5)
* **Sugar(nominal):** su - (0,1,2,3,4,5)
* **Red Blood Cells(nominal):** rbc - (normal,abnormal)
* **Pus Cell (nominal):**
* **Pus Cell clumps(nominal):** pcc - (present,notpresent)
* **Bacteria(nominal):** ba - (present,notpresent)
* **Blood Glucose Random(numerical):** bgr in mgs/dl
* **Blood Urea(numerical):** bu in mgs/dl
* **Serum Creatinine(numerical):** sc in mgs/dl
* **Sodium(numerical):** sod in mEq/L
* **Potassium(numerical):** pot in mEq/L
* **Hemoglobin(numerical):** hemo in gms
* **Packed Cell Volume(numerical):** pcv
* **White Blood Cell Count(numerical):** wc in cells/cumm
* **Red Blood Cell Count(numerical):** rc in millions/cmm
* **Hypertension(nominal):** htn - (yes,no)
* **Diabetes Mellitus(nominal):** dm - (yes,no)
* **Coronary Artery Disease(nominal):** cad - (yes,no)
* **Appetite(nominal):** appet - (good,poor)
* **Pedal Edema(nominal):** pe - (yes,no)
* **Anemia(nominal):** ane - (yes,no)
* **Class (nominal):** class - (yes,no)

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

Used pd.get\_dummies for converting categorical column (Nominal) to numerical column(One Hot Encoding)

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.

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.

1.Random\_Forest\_Classification:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s.no | criterion | Max\_features | N\_estimator | Cm Score |
| 1. | Log2 | Log\_loss | 100 | 98 |

2.Decision\_Tree\_Classification:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| s.no | criterion | Max\_features | Random state | splitter | Cm Score |
| 1. | gini | Log2 | None | random | 98 |

3.SVM:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s.no | c | gamma | kernel | Cm Score |
| 1. | 50 | scale | rbf | 69 |

4.Logistic\_Classification:

|  |  |  |  |
| --- | --- | --- | --- |
| s.no | penalty | solver | Cm score |
| 1. | L2 | Newton\_cg | 99 |

5.KNN:

|  |  |  |  |
| --- | --- | --- | --- |
| s.no | metrics | weight | Cm score |
| 1. | Minkowski | distance | 94 |

6.GaussianNB:

|  |  |  |  |
| --- | --- | --- | --- |
| s.no | priors | Var\_smoothing | Cm score |
| 1. | None | 0 | 98 |

7.Multinomial NB:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| s.no | alpha | Class\_prior | Fit\_prior | Force\_alpha | Cm Score |
| 1. | 1 | None | True | True | 82 |

8.BernouliiNB:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| s.no | alpha | Class\_prior | Fit\_prior | Force\_alpha | Binarize | Cm Score |
| 1. | 1 | None | True | True | 0.0 | 94 |

9.CategoricalNB:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s.no | alpha | Fit\_prior | Force\_alpha | Cm score |
| 1. | 0.1 | true | True | 99 |

10.complementaryNB:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| s.no | alpha | Fit\_prior | Force\_alpha | norm | Cm score |
| 1. | 0.1 | true | True | False | 84 |

6.The best Model is Logistic\_Classification which gave 99% which is performing good comparing to other model.