Chronic Kidney Disease Prediction:

**1.Problem statement:**

To create a predictive model which will predict chronic kidney disease (CKD) based on several parameters.

**2.) Information about the dataset (Total number of rows, columns)**

399 rows showing values of 399 different patients who were diagnosed CKD in past or not.

25 columns showing values of each factor that can contribute to chronic kidney disease.

**3.) Pre-processing method if any used (like converting string to number – nominal data)**

This dataset contains certain columns containing nominal values (rbc,pc,pcc,ba,htn,dm,cad,appet,pe,ane,classification (Disease diagnosed or not))

Hence Onehot encoding was used to modify the data from string to numerals.

**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.)**

| **Algorithms used** | **Classifier params** | **Overall accuracy** |
|---|---|---|
| Decision Tree Classifier | criterion = log_loss, max_depth 10, splitter random | 96% |
| Random Forest Classifier | criterion='log_loss', max_depth = 5,n_estimators=10 | 99% |
| Support Vector Classifier | C=1, kernel=linear | 98% |
| Logistic Regression | solver='liblinear', penalty = l2 | 98% |

| | | |
|---|---|---|
| K Nearest Neighbors | neighbors=3,weights=distance | 78% |
| Naïve Bayes | Gaussian<br>Multinomial<br>Complement<br>Bernoulli | 98%<br>81%<br>81%<br>93% |

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

The algorithm which worked best for this dataset is Random Forest Classifier with parameters criterion : log_loss, max_depth:5, n_estimators : 10