Assignment

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 .

Ans: It is supervised Machine learning .

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

Ans:399 rows and 24 columns.It is a unbalanced dataset.

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

Ans: 1. To change nominal data to categorical data

dataset=pd.get\_dummies(dataset,drop\_first=True)

2. standard scalar preprocessing

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

1. Develop a good model with good evaluation metric.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | F1\_score | Accuracy | ROC |
| Grid Random forest Classifier | 38% | 61% | 98% |
| Grid Decission Tree Classifier | 98% | 98% | 98% |
| Grid SVM | - | - | - |
| Grid Logistic Regression | 98% | 98% | 99% |
| KNN | 68% | 69% | 77% |
| Naive Bayes-MultinomialNB | 81% | 81% | 91% |
| Naive Bayes-BernoulliNB | 96%% | 96% | 98% |

1. .Best Model would be Grid Logistic Regression