

## **Classification Assignment**

### **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
- 2.) Tell basic info about the dataset (Total number of rows, columns)
- 3.) Mention the pre-processing method if you're doing any (like converting 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.
- 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.

### **SOLUTION:**

#### **1.) Identify your problem statement**

Stage 1: Supervised Learning

Stage 2: Apply Machine Learning

Stage 3: Classification Algorithm (because the target is categorical)

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

Total number of rows : 399

Total number of columns : 25

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

Solution:

Yes, I applied preprocessing to handle categorical (nominal) data. I used **one-hot encoding** to convert string values (like 'yes', 'no', 'normal', etc.) into numeric values (0 or 1). This was done using `pandas.get_dummies()` with `drop_first=True` to avoid the dummy variable trap. This transformation ensures that machine learning algorithms can work with categorical features in numeric form.

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.

Solution :

### **SVM GRID CLASSIFICATION ALGORITHM:**

### **RESULT OF SVM:**

**Roc.auc\_score value is : 1.0**

```
The F1 macro value for the best parameter{'C': 10, 'gamma': 'auto', 'kernel': 'poly'}: 1.0
```

The Confusion matrix:

```
[[51  0]
 [ 0 82]]
```

The report:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| False        | 1.00      | 1.00   | 1.00     | 51      |
| True         | 1.00      | 1.00   | 1.00     | 82      |
| accuracy     |           |        | 1.00     | 133     |
| macro avg    | 1.00      | 1.00   | 1.00     | 133     |
| weighted avg | 1.00      | 1.00   | 1.00     | 133     |

### **DECISION TREE GRID CLASSIFICATION:**

**Roc.auc\_score value is : 0.9817073170731707**

The F1 macro value for the best parameter{'criterion': 'gini', 'max\_features': None, 'splitter': 'random'}:  
0.9775556904684072

The Confusion matrix:  
[[51 0]  
[ 3 79]]

The report:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| False        | 0.94      | 1.00   | 0.97     | 51      |
| True         | 1.00      | 0.96   | 0.98     | 82      |
| accuracy     |           |        | 0.98     | 133     |
| macro avg    | 0.97      | 0.98   | 0.98     | 133     |
| weighted avg | 0.98      | 0.98   | 0.98     | 133     |

**BEST PAPAMETER:**

|   |          |          |          |          |      |      |        |  |          |          |
|---|----------|----------|----------|----------|------|------|--------|--|----------|----------|
| 1 | 0.004226 | 0.001195 | 0.004207 | 0.001345 | gini | None | random | {'criterion':<br>'gini',<br>'max_features':<br>None, 's... | 0.981569 | 0.962264 |
|---|----------|----------|----------|----------|------|------|--------|--|----------|----------|

**RANDOM FOREST GRID CLASSIFICATION:**

The F1 macro value for the best parameter{'criterion': 'gini', 'max\_features': 'sqrt', 'n\_estimators': 100}:  
0.9849624060150376

The Confusion matrix:  
[[50 1]  
[ 1 81]]

The report:

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| False        | 0.98      | 0.98   | 0.98     | 51      |
| True         | 0.99      | 0.99   | 0.99     | 82      |
| accuracy     |           |        | 0.98     | 133     |
| macro avg    | 0.98      | 0.98   | 0.98     | 133     |
| weighted avg | 0.98      | 0.98   | 0.98     | 133     |

**Final Model Selection:**

**The Support Vector Machine (SVM) model was selected as the final model because it achieved perfect performance on the test dataset. The key evaluation metrics — Accuracy, Precision, Recall, F1 Score, and ROC-AUC Score — all had values of 1.0, indicating a perfect classification.**

**Therefore, the SVM model is considered the best fit for this dataset.**