**Assignment – Classification Algorithm**

**Problem Statement:**

Hospital management wants to predict the Chronic Kidney Disease (CKD) based on several parameters.

**Problem Identification:**

Machine Learning - Supervised – Classification

**Basic Info about dataset:**

The dataset consist of 25 Columns and 399 Rows.

**Output data:** Classification Column

This dataset consist of 10 nominal data columns, here we have to convert string to numbers for creating best model

**Machine Learning Classification**

**Phase 01:**

1. **Logistic Regression**

The f1\_macro value for best parameter {'penalty': 'l2', 'solver': 'sag'}: 0.9924946382275899

The confusion Matrix:

[[51 0]

[ 1 81]]

The report:

precision recall f1-score support

0 0.98 1.00 0.99 51

1 1.00 0.99 0.99 82

accuracy 0.99 133

macro avg 0.99 0.99 0.99 133

weighted avg 0.99 0.99 0.99 133

roc\_auc\_score = 1.0

1. **Support Vector Machine (SVM)**

The f1\_macro value for best parameter {'C': 10, 'gamma': 'scale', 'kernel': 'sigmoid'}: 0.9924946382275899

The confusion Matrix:

[[51 0]

[ 1 81]]

The report:

precision recall f1-score support

0 0.98 1.00 0.99 51

1 1.00 0.99 0.99 82

accuracy 0.99 133

macro avg 0.99 0.99 0.99 133

weighted avg 0.99 0.99 0.99 133

roc\_auc\_score = 1.0

1. **Decision Tree**

The f1\_macro value for best parameter {'criterion': 'entropy', 'max\_features': 'sqrt', 'splitter': 'random'}: 0.9626932787797391

The confusion Matrix:

[[51 0]

[ 5 77]]

The report:

precision recall f1-score support

0 0.91 1.00 0.95 51

1 1.00 0.94 0.97 82

accuracy 0.96 133

macro avg 0.96 0.97 0.96 133

weighted avg 0.97 0.96 0.96 133

roc\_auc\_score = 0.9695121951219512

1. **Random Forest**

The f1\_macro value for best parameter {'criterion': 'entropy', 'max\_features': 'log2', 'n\_estimators': 100}: 0.9849624

The confusion Matrix:

[[50 1]

[ 1 81]]

The report:

precision recall f1-score support

0 0.98 0.98 0.98 51

1 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

roc\_auc\_score = 0.99976

1. **KNN**

The f1\_macro value for best parameter {'algorithm': 'auto', 'weights': 'distance'}: 0.9404945931261721

The confusion Matrix:

[[51 0]

[ 8 74]]

The report:

precision recall f1-score support

0 0.86 1.00 0.93 51

1 1.00 0.90 0.95 82

accuracy 0.94 133

macro avg 0.93 0.95 0.94 133

weighted avg 0.95 0.94 0.94 133

roc\_auc\_score = 1.0

1. **Naive bayes**

**GaussianNB**

The confusion Matrix:

[[51 0]

[ 3 79]]

The report:

precision recall f1-score support

0 0.94 1.00 0.97 51

1 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

**MultinomialNB**

The confusion Matrix:

[[51 0]

[ 2 80]]

The report:

precision recall f1-score support

0 0.96 1.00 0.98 51

1 1.00 0.98 0.99 82

accuracy 0.98 133

macro avg 0.98 0.99 0.98 133

weighted avg 0.99 0.98 0.99 133

**BernoulliNB**

The confusion Matrix:

[[51 0]

[ 7 75]]

The report:

precision recall f1-score support

0 0.88 1.00 0.94 51

1 1.00 0.91 0.96 82

accuracy 0.95 133

macro avg 0.94 0.96 0.95 133

weighted avg 0.95 0.95 0.95 133

**ComplementNB**

The confusion Matrix:

[[51 0]

[ 2 80]]

The report:

precision recall f1-score support

0 0.96 1.00 0.98 51

1 1.00 0.98 0.99 82

accuracy 0.98 133

macro avg 0.98 0.99 0.98 133

weighted avg 0.99 0.98 0.99 133

**Phase 02:**

**Best Model Selection:**

**Both SVM and Logistic Regression model shows accuracy – 0.99**

The **f1\_macro value** for best parameter {'penalty': 'l2', 'solver': 'sag'}: 0.9924946382275899

**The confusion Matrix:**

[[51 0]

[ 1 81]]

**The report:**

precision recall f1-score support

0 0.98 1.00 0.99 51

1 1.00 0.99 0.99 82

accuracy 0.99 133

macro avg 0.99 0.99 0.99 133

weighted avg 0.99 0.99 0.99 133

**roc\_auc\_score** = 1.0