

## **Classification Assignment**

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

The goal of this project is to build a predictive model that can classify whether a patient is suffering from Chronic Kidney Disease (CKD) based on a set of clinical parameters provided by the hospital.

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

- **Total number of rows:** 399
- **Total number of columns:** 28

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

- `pd.get_dummies()` converts **categorical string columns** into multiple **binary columns** (one-hot encoding).
- The parameter `drop_first=True` drops the first category to avoid dummy variable trap (redundant columns).

#### **· Feature Scaling:**

Numerical features were standardized using **StandardScaler** to ensure features are on a similar scale, which helps some algorithms (like SVM or KNN) perform better.

#### **· Train-Test Split:**

The dataset was split into training and testing subsets to evaluate the model performance on unseen 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.**

SVC model's report:

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

- **Precision** of 1.00 indicates the model never misclassified negatives as positives, and vice versa.
- **Recall** of 1.00 means the model identified all positive and negative cases correctly.
- **F1-score** of 1.00 reflects a perfect balance between precision and recall.

**5.) All the research values of each algorithm should be documented.**

**(You**

**can make tabulation or screenshot of the results.)**

**SCV:**

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:**

The report:

	precision	recall	f1-score	support
False	0.83	0.98	0.90	51
True	0.99	0.88	0.93	82
accuracy			0.92	133

macro avg	0.91	0.93	0.91	133
weighted avg	0.93	0.92	0.92	133

### Random Forest:

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

### Logistic Regression:

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

### KNN:

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

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

Both **Support Vector Classifier (SVC)** and **Logistic Regression** achieved perfect evaluation metrics (precision, recall, f1-score all equal to 1.00) on the test set, indicating excellent performance for both models.

However, **SVC was selected as the final model**