Chronic Kidney Disease prediction using AI

1.1Background

Chronic Kidney Disease (CKD) is a medical condition characterized by the gradual loss of kidney function over time. ML (Machine Learning) classification models can be employed to predict the presence or absence of CKD based on relevant patient data. Here's a short overview of how ML can be used for CKD classification:

1.2 Data Collection:

Features: Gather relevant patient data, including demographic information, medical history, laboratory test results (e.g., serum creatinine levels), and other clinical indicators.

Labels: Assign labels indicating the presence or absence of CKD based on diagnostic criteria.

1.3 Data Preprocessing:

Cleaning: Handle missing values, outliers, and any data inconsistencies.

Normalization/Standardization: Ensure that features are on a consistent scale.

Encoding: Convert categorical variables into a format suitable for ML algorithms.

1.4 Model Selection:

Choose a suitable ML classification algorithm. Common choices include Decision Trees, Random Forests, Support Vector Machines, and Neural Networks.

1.5 Training the Model:

Split the Data: Divide the dataset into training and testing sets to assess model generalization.

Model Training: Train the chosen algorithm on the training data.

1.6 Model Evaluation:

Testing: Use the trained model to predict CKD status on the testing set.

Performance Metrics: Evaluate the model using metrics such as accuracy, precision, recall, F1 score, and confusion matrix.

1.7 Model and Validation:

Interpret the model's predictions to understand the factors contributing to CKD classification.

Validate the model using additional datasets, if available, to ensure robustness.

1.8 Continuous Monitoring and Improvement:

Continuously monitor the model's performance in real-world scenarios.

Update the model periodically based on new data or changes in medical understanding.

Using ML for CKD classification contributes to early detection and intervention, potentially improving patient outcomes. However, it's crucial to collaborate with healthcare professionals and ensure that

the model aligns with medical guidelines and ethical considerations. Additionally, interpretability and transparency in model predictions are vital for gaining trust from healthcare practitioners and patients.

Solution:

After analyse the customers data, we have to predict the CKD.

1.) Identify your problem statement

- We have clear set of data's with numbers so will take ML

Domain Selection- Machine Learning

- We have clear requirements to predict the CKD

Learning selection- Supervised learning

- The requirements for the prediction is numeric value.

Departments – Classified

1.) Tell basic info about the dataset

Columns – 28
Rows - 399

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

Since the values are in words we converting words to number using One Hot Encoding by nominal data

Logistic Regression

	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

Decision Tree

	precision	recall	f1-score	support
0	0.94	0.96	0.95	51
1	0.98	0.96	0.97	82
accuracy			0.96	133
macro avg	0.96	0.96	0.96	133
weighted avg	0.96	0.96	0.96	133

SCM

	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

RF

	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

in this model we got accuracy result of almost 0.99 for all models.