**INTRODUCTION**

**OVERVIEW**

In today's era everyone is trying to be conscious about health although due to workload and busy schedule one gives attention to the health when it shows any symptoms of some kind. But CKD is a disease which doesn't shows symptoms at all or in some cases it doesn't show any disease specific symptoms it is hard to predict, detect and prevent such a disease and this could be lead to permanently health damage, but machine learning can be hope in this problem it is best in prediction and analysis. By using data of CKD patients with 14 attributes and 400 record we are going to use various machine learning techniques like Decision Tree, linear regression, etc. to build a model with maximum accuracy of predicting whether CKD or not and if yes then its Severity.

**PURPOSE**

Kidney function tests are urine or blood tests that evaluate how well your kidneys are working.

Some conditions, such as [diabetes](https://my.clevelandclinic.org/health/diseases/7104-diabetes-mellitus-an-overview) or [high blood pressure (hypertension)](https://my.clevelandclinic.org/health/diseases/4314-hypertension-high-blood-pressure), affect how well the kidneys work. If you have one of these conditions, your healthcare provider may use kidney function tests to help monitor these conditions.

So to identify the quicker results regarding the tests, means quicker medications, so this analysis will be helpful for us to test the chronic kidney disease

**EXISTING PROBLEM**

CKD is a disease which doesn't shows symptoms at all or in some cases it doesn't show any disease specific symptoms it is hard to predict, detect and prevent such a disease and this could be lead to permanently health damage

**PROPOSED SOLUTION**

Machine learning can be hope in this problem it is best in prediction and analysis

By using data of CKD patients record we are going to use various machine learning techniques like Decision Tree, linear regression, etc. to build a model with maximum accuracy of predicting whether CKD or not and if yes then its Severity.

**EXPERIMENTAL INVESTIGATION**

In this paper CKD dataset Is downloaded from KAGGLE repository. This dataset includes 400 patients’ records with 25 attributes. All this 25 attributes are main attributes which are related to CKD disease. Out of 25 attributes we only use 9 attributes to build our predictive model.

DATASET -

Dataset of prediction of chronic kidney disease using machine learning algorithm is downloaded from UCI repository

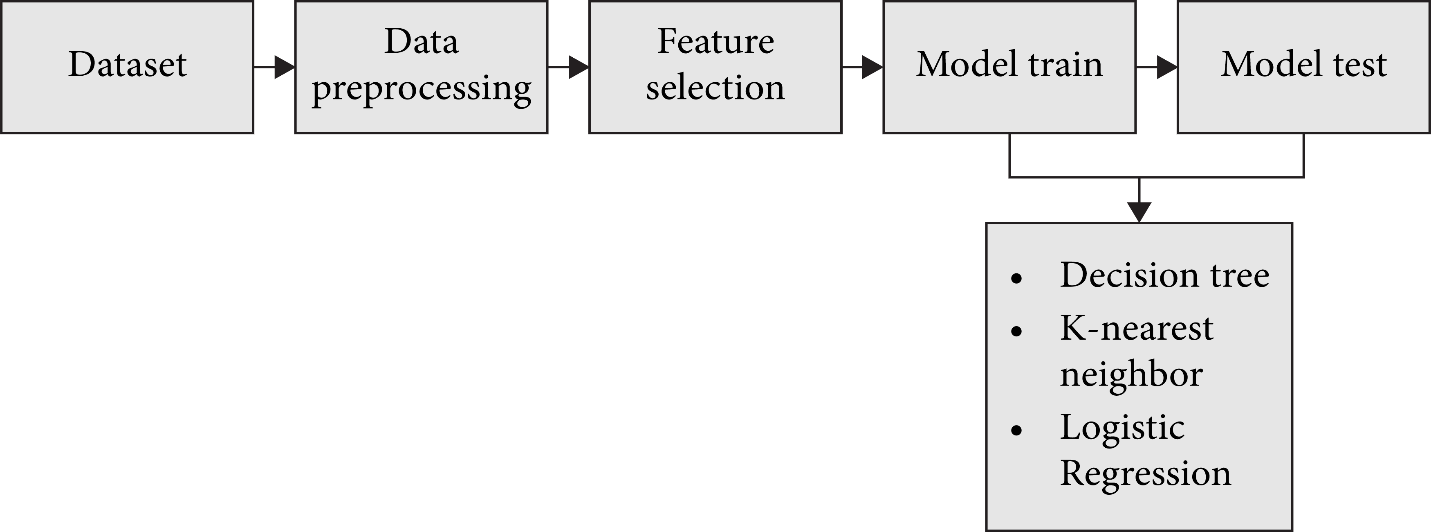
DATA PREPROCESSING:

Data Cleaning: Gather open source raw data of CKD patients available on internet. Data obtained from internet does not contains the name of the attribute so first we assigned the names to the attribute. Missing values in the dataset like NA’s or blank values are removed by using “ReplaceMissingValues”, which replaces NA’s with the mean and mode values of that attribute

CKD PREDICTION USING MACHINE LEARNING MODELS

The paper tries to propose a data mining framework for knowledge discovery on the CKD datasets. Large amounts of CKD datasets are collected. Data preparation and preprocessing is done using the traditional methods of data mining process.

**FLOW CHART**



**Algorithm:**

Input: Chronic Kidney Disease Dataset

Output: High Accuracy prediction Framework

Step1: Input data

Step2: Pre-process the data

Step 2.1: Convert Categorical values to numerical values

Step 2.2: Replace numerical missing values by Mean

Step2.3: Replace Categorical missing values by Mode

Step3: Construct Classifier Models

Step3.1: Construct the deserved model by testing and training

Step 4: Check the accuracy of the constructed models using confusion matrix.

Step 5: Now create a pkl file to address the model in the flask

Step 6: Create a HTML code for analysis

Step 7: Develop the flask code which links to the HTML web page and create a app.py

Step 8: Now open the link localhost:5000 and predict

**RESULTS AND DISCUSSION**

Models has been constructed using training data set(280 instances) which is 70% of original CKD data set. Constructed models have been validated using test data which is 30% of original data with respect to the parameter accuracy. Here, Accuracy has been calculated using confusion matrix .The best classifier model is the one with highest accuracy..

**ADVANTAGES AND DISADVANTAGES**

Early CKD prediction could lead to therapeutic interventions and lifestyle changes, prevention of progression to higher stages, and reduction of dependency as well as costly healthcare spending

However, there remains a need for kidney disease prediction for patients newly diagnosed with T2DM who are at high risk of CKD development

Additionally, early intervention could significantly improve patient quality of life as patients Emerging technologies, pharmacology and therapeutics with CKD report disease and management affecting not only their physical health, but also mental and social health

**CONCLUSION AND FUTURE SCOPE**

This study investigates various machine learning techniques, particularly classification and association techniques, to predict CKD. The study analyzes the effects of using feature selection techniques in combination with classification techniques. The results are compared for correctly classified instances, and mean absolute value with and without the feature selection technique. The results note that the best result can be achieved with the Apriori associative algorithm for 97% accuracy.

Future research should analyze different supervised and unsupervised machine learning techniques and feature selection techniques with additional performance metrics for better CKD prediction.

**SOURCE CODE**

https://github.com/smartinternz02/SI-GuidedProject-8286-1642769710