**DIAGNOSE KIDNEY DISEASE**

**INTRODUCTION**

Data mining is the process of extracting hidden knowledge from large volumes of data. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Data mining plays an efficient role in prediction of diseases in health care industry.

Prediction of diseases by analysis of voluminous historical data is one of the most significant applications of data mining. Medical data mining is the process of finding useful patterns that would be helpful in medical diagnosis. The predictability of CHRONIC KIDNEY DISEASE will be more effective since earlier detection of disease will be helpful for the patients to take care of themselves.

**PROBLEM STATEMENT**

**“Diagnose about Chronic Kidney Disease from Medical Test”**

Chronic diseases have become a major cause of global morbidity and mortality, 4 out of 5 chronic disease deaths now occur in low- and middle-income countries. In India the projected number of deaths due to chronic diseases will rise from 3.78 million in 1990 (40.4% of all deaths) to an expected 7.63 million in 2020 (66.7% of all deaths)

Chronic kidney disease (CKD) can be diagnosed with blood and urine tests. In many cases, it's only picked up because a routine blood or urine test indicates that the kidneys may not be working normally.

**SCOPE**

Our examination concentrates on this part of Medical conclusion learning design through the gathered data of Chronic Kidney Disease and to create smart therapeutic choice emotionally supportive network to help the physicians.

**PROPOSED WORK**

In the proposed system data mining has been done in different phases .Data set has to be divided into different training and testing dataset.

**Data Pre-processing**:

In this phase dataset has been pre-processed for the removal of empty sets and anomalies available in the dataset. Data cleaning removes the observations containing noise and those missing data.

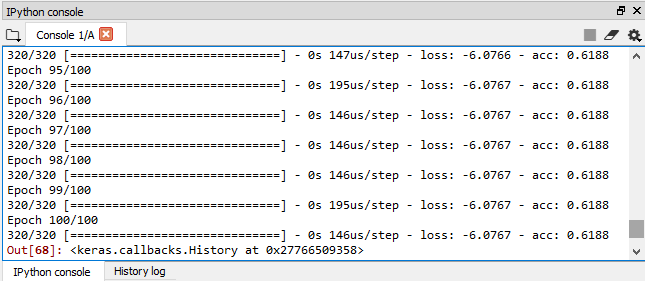
**IMPLEMENTATION USING NEURAL NETWORK**

NNs are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value. Each link is associated with weight.

**Steps:**

* The processed data is splitted into features and labels.
* Then it will be divided into train and test sets.
* Then the independent variable will be feed to the feature scaling to distribute the data normally.
* Neural Network algorithm is implemented by using keras, tensorflow, theano.
* The accuracy is noted.

**OUTPUT**



**The Accuracy is about 61%**

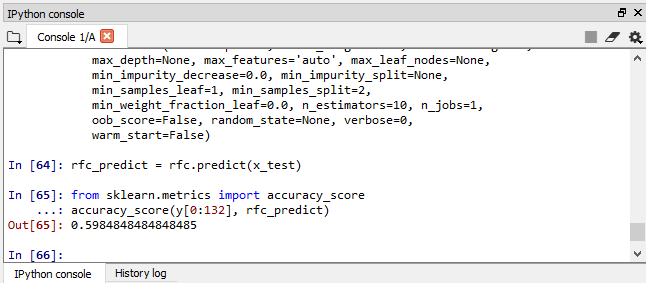
**IMPLEMENTATION USING RANDOM FOREST ALGORITHM**

As the name suggest, this algorithm creates the forest with a number of trees. The higher the number of trees gives the high accuracy results. Several decision trees combined to form a single random forest. Random forest can be used as both regression and classification tasks. Random forest classifiers can handle missing values.

**Steps:**

* Data processing plays a vital role in deploying the algorithm.
* Then it will be splitted into train and test sets.
* Applying random forest classifier algorithm in it.
* Prediction will be based on the features and the accuracy is monitored.

**OUTPUT**

****

**The Accuracy is about 59% (which is lesser than using NN)**

**FUTURE WORK**

Our ongoing process is to develop a web application which is used for the normal persons to access the system as easier. This can be done by PYTHON FLASK which is a web service that has no database abstraction layer, form validation or any other components. Flask is considered more pythonic because in common situations flask web application is more explicit.

**CONCLUSION**

The model has been trained to predict the risk of chronic kidney disease. The performance of the model has been evaluated using the performance measures such as accuracy, sensitivity and specificity. The performance of the algorithm has been improved by feature subset selection and by varying the size of the training dataset. The experiment result proves that **ANN (Artificial Neural Network) provides the highest accuracy than other techniques.**