

MACHINE LEARNING PROJECT :

Chronic Kidney Disease Prediction using Artificial Neural Network

Nidhi Jaiswal

Student ID: 1823025

LY/EXTC

COURSE : Data Science Platter

Introduction:

Kidney Chronic Disease is a condition characterised by a gradual loss of kidney function.

Common (More than 10 lakh cases per year in India)

Data was collected for this cross-sectional study in a standardized format according to predetermined criteria. Of the 52,273 adult patients, 35.5%, 27.9%, 25.6% and 11% patients came from South, North, West and East zones respectively.

Early stages show no symptoms. When the disease progresses symptoms presented are:

- Nausea
- Loss of appetite
- Fatigue
- Oedema- swelling of feet and ankle
- Need to urinate more often, especially at night
- High blood pressure (hypertension)
- Decreased mental sharpness
- Problems with sleep
- Blood in urine
- Protein in urine

Predictions are made from measuring several parameters in the body taking symptom into consideration.

2. Problem definition

2.1 Task definition

Predicting the chronic kidney disease by taking several parameters into consideration like diabetes mellitus, appetite, blood test, etc. All these features are considered as input and the label with output ckd or nonckd. And this analysis will help us to understand this concept effectively.

2.2 Methodology

1. Binary classification is used here to classify the entities.
2. The Keras Library that comes along with tensorflow library is employed to generate the Deep learning model.
3. The dataset is used from Kaggle and One hot encoder is used for data simplification and arrangement.

4. Then the data is splitted into independent features and dependent vector before feeding it to neural network.
5. The data is divided into test and train set which is achieved using `test_train_split` function provided in the `model_selection` class of `sklearn` module.
6. The splitted data is such that 70% randomly selected data is put into train test and rest of the 30% for test set/
7. The neural network that has three layers .There are two layers of nodes and one output node .The last node uses the sigmoid activation function that will squeeze all the value between 0 and 1 into sigmoid curve . The other two layers use ReLU as the activation function.
8. It uses Adam a momentum-based optimizer and the loss function used is `binary_crossentropy`
9. The confusion matrix , accuracy and `f1_score=1`

Conclusion :

A working Neural Network for the binary classification is developed .The same problem can be resolved using other algorithms such as Logistic Regression , Navie Bayes etc. Deep network or neural network are generally recommended if the data size is large.