PROJECT REPORT

Early Prediction Of Chronic Kidney Disease Using IBM Watson Studio

1. INTRODUCTION

1.1 Overview

Machine learning has earned a remarkable position in the healthcare sector because of its capability to enhance disease prediction in the healthcare sector. Artificial intelligence and Machine learning techniques are being used in the healthcare sector. Nowadays, one of the world's crucial health-related problem is kidney disease. It is increasing day by day because of not maintaining proper food habits, drinking less amount of water and lack of health consciousness.

Chronic Kidney disease can also be termed as kidney failure. The present work focuses on predicting weather a person is suffering from CKD or not using Data mining using Machine Learning. Data Mining is the process of searching large data sets and discovering patterns and trends and transforming it into understandable data using Data pre-processing, Visualization. Machine Learning is the field of computer which uses statistical techniques to give the ability to learn to computer. Machine learning can be both Supervised and Unsupervised learning. Supervised learning can be defined as when we map an input to a desired output. Machine learning algorithms are provided to support future predictions.supervised machine learning algorithms like Logistic regression, multi-class classification, support vector machine, K-nearest neighbour, Naïve Bayes, Random Forest and many more. In un-Supervised learning algorithm, we train the data using info which is not labelled. In thisalgorithm we divide the data into 2 groups based on similarity and reducing the dimensionality. Most common unsupervised learning approaches are clustering algorithms. There are various clustering algorithms like Hierarchical clustering, K-means clustering and many more. Feature Selection also called as variable selection, attribute selection or feature extraction. It uses relevant data sets and avoids redundant and irrelevant data. It reduces the dimensionality by using small subsets from the original dataset it helps in easy calculation of results and attaining shorter training times. In this research we have used machine learning to detect CKD and Non-CKD by using 25 attributed which Contribute to kidney disease. The data used consist of record of 400 people. The data set has various missing data. We have used this data sets and classification algorithms to build a classification model for prediction of CKD. The model with the best accuracy prediction is taken. This will help to achieve fast and accurate results for CKD prediction, which will reduce the time for disease prediction and provide benefits to both doctors and patients in providing early treatment and speedy recovery.

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated in the early stages. Usually, people are not aware that medical tests, we take for different purposes could contain valuable information concerning kidney diseases. Consequently, attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease. The information says that it helps us to measure the severity of the problem and we make use of such information to build a machine learning model that predicts Chronic Kidney Disease.

1.2 Purpose

The main goal of this project is to predict whether an individual will have chronic kidney disease or not based on the data provided.

2. LITERATURE SURVEY

2.1 Existing problem

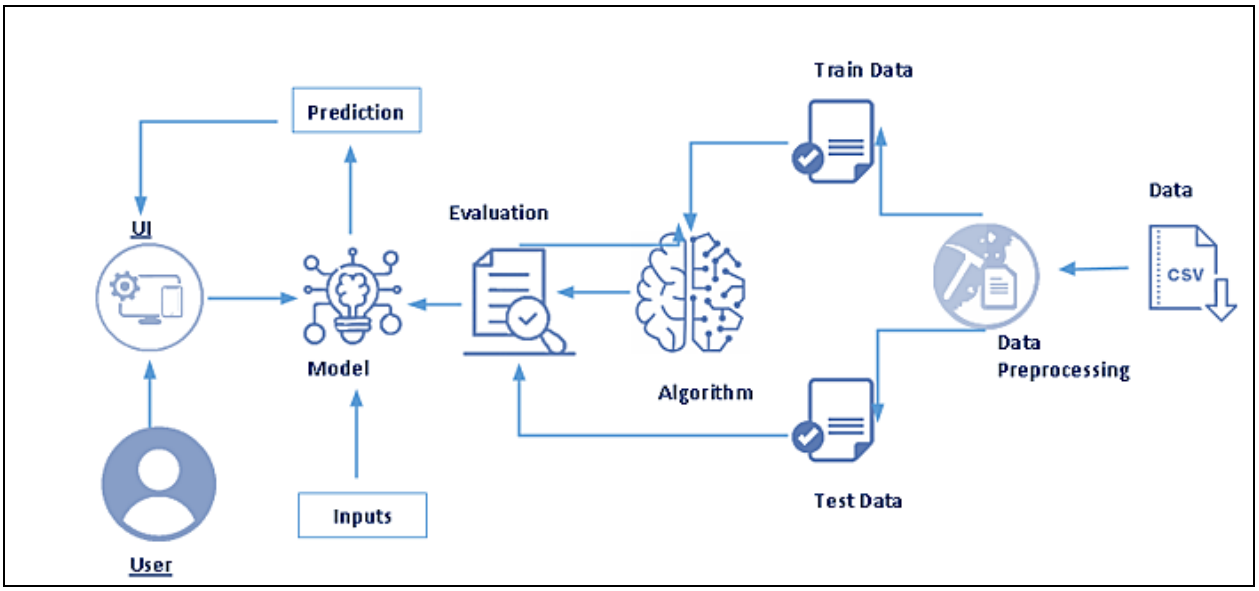
- In the current existing system so many researchers working with data mining algorithms in different kidney disease survey techniques.
- In these survey techniques they are using different databases available for different sources, but they are unable to analyze data with visualization techniques to identify the correlation between different attributes.
- Using limited techniques and they are unable to optimize by increasing efficiency.

2.2Proposed solution

In this proposed system we are able to identify the patients with disease. Once any person gets kidney disease, they may suffer from the disease which may decrease their working capability as well as living quality. Our aim is to predict patients with chronic kidney failure (ckd) disease and patients who do not (not-ckd) suffer from the disease. So, for that we are building a Machine Learning model to predict the compressive strength of concrete using IBM Watson studio MachineLearning Service. The model is deployed on IBM cloud to get a scoring end point which can be used as web applications. We are developing a web application which is built using Flask service. We make use of the scoring end point to give user input values to the deployed model. The model prediction is then showcased on User Interface.

3. THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

Hardware: Windows 10

Software:

Anaconda Navigator :

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS. Conda is an open-source, cross-platform, package management system. Anaconda comes with so very nice tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using **Jupyter** notebook and **Spyder**.

To build Machine learning models you must require the following packages,

- **Sklearn:** Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.

- **NumPy:** NumPy is a Python package that stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object

- **Pandas:** pandas is a fast, powerful, flexible, and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

- **Matplotlib:** It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits

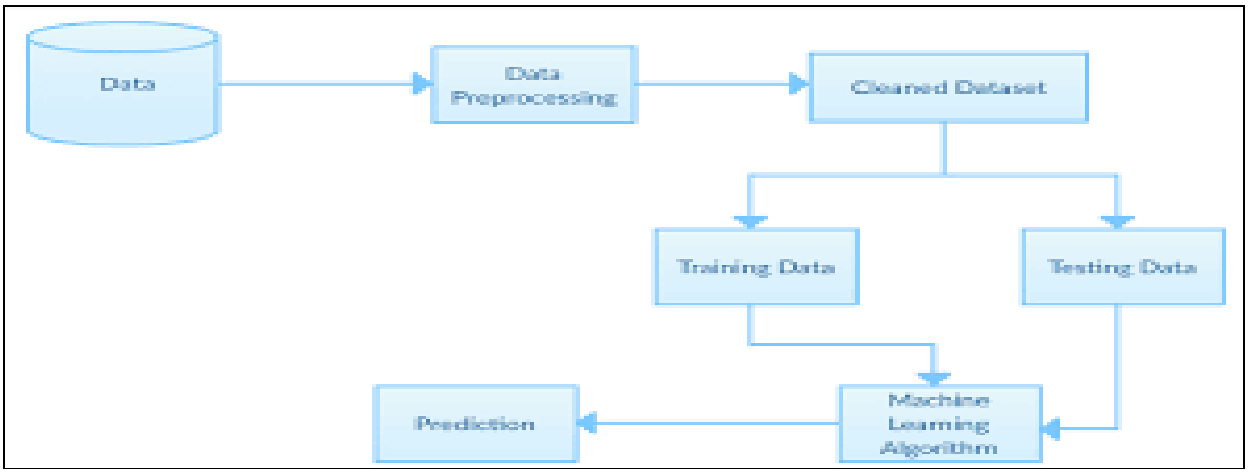
- **Flask:** Web framework used for building Web applications.

4. EXPERIMENTAL INVESTIGATIONS

The dataset consists of 400 instances with 25 attributes and has no missing values. There are 8 input variables and 1 output variable. Two input variables represent the amount of blood urea and blood glucose random and six input variables representing the different symptoms.

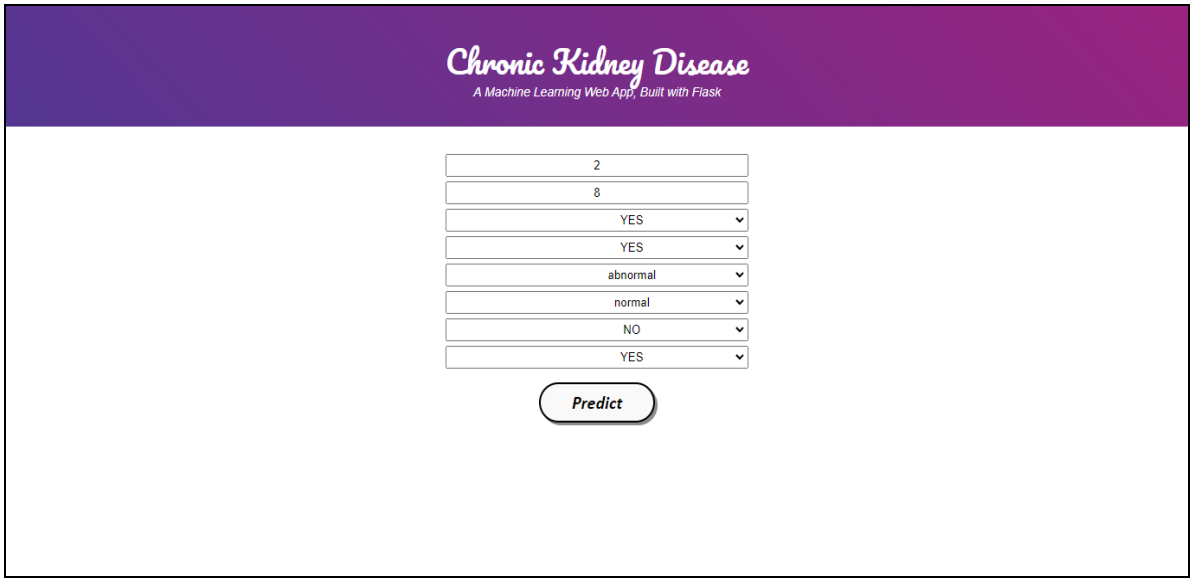
We shall explore the data to see how input features are affecting compressive strength. The first step in a Data Science project is to understand the data and gain insights from the data before doing any modelling. This includes checking for any missing values, plotting the features with respect to the prediction, observing the distributions of all the features, and so on. Let us import the data and start analysing it.

5. FLOWCHART



6. RESULT

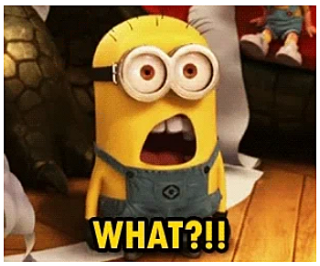
After the testing of the project as given below, we created web pages. Hear predicts whether he or she was affected by chronic kidney disease or not. The flask is used web application to help the people to be aware of the health situation.



Chronic Kidney Disease

A Machine Learning Web App, Built with Flask

Prediction: Oops! You have Chronic Kidney Disease.

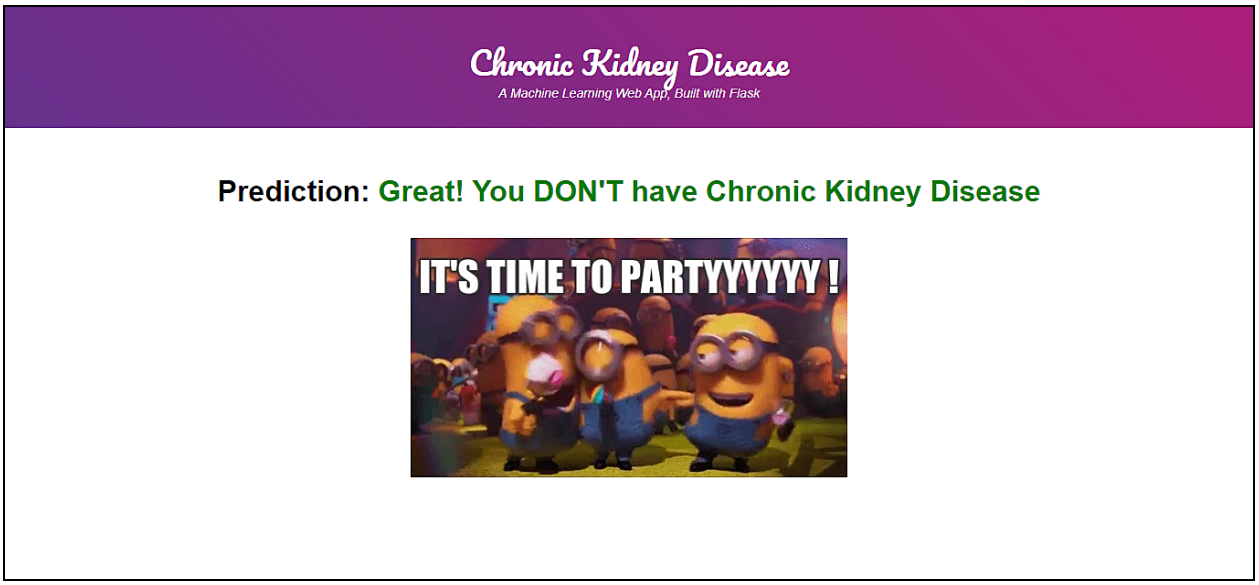


Chronic Kidney Disease

A Machine Learning Web App, Built with Flask

-12	
-4	
YES	
NO	
normal	
abnormal	
NO	
NO	

Predict



7. ADVANTAGES & DISADVANTAGES

Advantages:

- The prediction gives good insights about the risk of kidney disease in the body.
- With the help of this UI, Efficient prediction of kidney disease that can be done in an easy way.

Disadvantages:

- The model may need to be re-trained in case of decrementation of patients.
- Many times we do face a situation where we find an imbalance in data which leads to poor accuracy of models.

8. APPLICATIONS

Using The IBM Watson studio and Flask, we can build and deploy a machine learning model with sophisticated training features and no coding. The tool does most of the work for us. In this project, the UI model building can help people a lot. If we use machine learning then we can predict if a person may get the diseases in future, then we can give treatment accordingly by the help of machine learning we can save the lives of the people. The Flask service provides us a better user UI with the help of anyone who can deploy machine learning models and get predicted results.

9. CONCLUSION

In this project we have discussed the direct impact of machine learning on health systems, but have not explored the indirect effects of machine learning in basic sciences, drug discovery and other enabling technologies on health systems. Prediction is inherently difficult: technology modifies its environment and the environment then generate further opportunities and new constraints for the technology. Ultimately, general purpose intelligence will be possible, as a version of it already exists in human brains. However, an extrapolation of existing techniques to re-create general intelligence artificially appears unlikely in the next 5-10 years. However, what is immediately plausible, and should therefore be planned for, is a federation of ‘narrow’ and ‘targeted’ machine learning systems that are able to tackle core information processing problems across a health system by augmenting capabilities of human decision makers, and in so doing establishing new standards of effectiveness and efficiency in clinical and management operations. This is a significant opportunity for health system transformation as the cost of augmenting decision-making capabilities.

10. BIBLIOGRAPHY

http://www.ijamtes.org/gallery/41.%20may%20ijmte%20-%20490.pdf

smartinternz.com