

Early Detection of Chronic Kidney Disease using Machine Learning

Team ID : PNT2022TMID25893

Team Size : 4

Team Leader : SAM JABEZ S

Team member : SURYA S

Team member : SHRIVARSHAN C

Team member : SYED FARHAN A

INTRODUCTION

Project Overview: -

Chronic kidney disease prediction is one of the most important issues in healthcare-analytics. The most interesting and challenging tasks in day-to-day lives as one third of the adult population is affected by chronic kidney disease (CKD), and millions die each year because they do not have access to affordable treatment. Chronic Kidney Disease can be cured, if treated in the early stages. The main aim of the project is to predict whether the patient have chronic kidney disease or not in a painless, accurate and faster way based on certain diagnostic measurement like Blood Pressure(BP), Albumin(AI) etc., and then appropriate treatment can be given based on the details provided by the model.

Purpose: -

The purpose of the project is to alert doctors for an early detection of kidney disease and hence ensure speedy recovery or prevention of kidney disease. This Project aims at creating a model for early detection of Chronic Kidney Disease using Machine Learning technology. The output is integrated with Flask. The front end developed in html is used to receive user input on various parameters needed to decide on the early detection of kidney disease. The same model is deployed into IBM cloud using API keys and scoring endpoints.

LITERATURE SURVEY

Existing Problem:-

Presently kidney disease is detected at late stages in many countries leading to loss of precious lives. There are very few means to identify them at an early stage. Most of the user details remain unverified and it's difficult to track the fake users. The user interface of the application is not user friendly and the user must have a device with an android operating system with an active internet connection to interact with this application

LITERATURE SURVEY

Q.-L. Zhang & D. Rothenbacher [1] proposed a machine learning methodology for diagnosing CKD. Chronic kidney disease is becoming a major public health problem worldwide. This article reviews the published evidence of prevalence of CKD in population-based study samples that used the standardized definition from the kidney disease outcomes quality initiative of the National Kidney Foundation (K/DOQI) practice guideline and particularly focus on performance of serum-creatinine based equations for GFR estimation.

W. M. McClellan et. al [2], "Albumin-uria and racial disparities in the risk for ESRD proposed a machine learning methodology for diagnosing CKD. Socio Economic status (SES) is independently associated with chronic kidney disease progression however the association with other CKD outcomes is unclear. In particular, the potential differential effect of SES on mortality among blacks and whites is understudied in CKD.

M. K. Haroun [3] said that chronic kidney disease is an increasing cause of morbidity and mortality in the United States. Prospective data on risk factors for CKD are limited to men, and few studies examine the importance of smoking. All cases were confirmed as CKD by medical chart review. Adjusted relative hazards of CKD were modeled using Cox proportional hazards regression including age as the time variable and baseline BP, cigarette smoking, gender, and diabetes status as risk factors. The adjusted hazard ratio (95% confidence interval) of developing CKD among women was 2.5 (0.05 to 12.0) for normal BP, 3.0 (0.6 to 14.4) for high-normal BP, 3.8 (0.8 to 17.2) for stage 1 hypertension, 6.3 (1.3 to 29.0) for stage 2 hypertension, and 8.8 (1.8 to 43.0) for stages 3 or 4 hypertension compared

with individuals with optimal BP.

W. D. Souza et. al [4], said that Chronic Kidney Disease is one of the most critical illness nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. The important feature selection technique was also applied to the dataset. For each classifier, the results have been computed based on full features, correlation-based feature selection, Wrapper method feature selection, Least absolute shrinkage and selection operator regression, synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features.

W. Mula-Abed et. al [5], explained that chronic kidney disease is an important epidemic and public health problem that is associated with a significant risk for vascular disease and early cardiovascular mortality as well as progression of kidney disease. Currently it is classified into five stages based on the glomerular filtration rate as recommended by many professional guidelines. Radio labelled methods for measuring GFR are accurate but not practical and can be used only on a very limited scale while the traditional methods require timed urine collection with its drawback of inaccuracy, cumbersomeness and inconvenience for the patients.

A. S. Levey et. al [6], explained that Changes in proteinuria have been suggested as a surrogate outcome for kidney disease progression to facilitate the conduct of clinical trials. This report summarizes a workshop sponsored by the National Kidney Foundation and US Food and Drug Administration with the following goals to evaluate the strengths and limitations of criteria for assessment of proteinuria as a potential surrogate end point for clinical trials in chronic kidney disease to explore the strengths and limitations of available data for proteinuria as a potential surrogate end point and to delineate what more needs to be done to evaluate proteinuria as a potential surrogate end point.

S. Gerogianni [7] explained that Chronic Renal Failure (CRF) is a public health problem that has serious impact on mental and psychological health of patients undergoing haemodialysis. The aim of this quantitative research study was to investigate the psychological impact of dialysis on the quality of life of patients with End Stage Renal Failure (ESRF). The sample study included 100 patients undergoing haemodialysis in four hospitals in Athens. Data was collected by the completion of a questionnaire KDQOL-SF, incorporating

the tool overview of the SF-36 Health and an additional questionnaire that included demographic characteristics.

K. R. A. Padmanaban & G. Parthiban [8] proposed a machine learning methodology for diagnosing CKD. This paper aims at predicting the early detection of chronic kidney disease also known as chronic renal disease for diabetic patients with the help of machine learning methods and finally suggests a decision tree to arrive at concrete results with desirable accuracy by measuring its performance to its specification and sensitiveness. On comparing the classification algorithms with respect to Naive Bayes and Decision tree, we came to conclusion that the accuracy is up to 91% for Decision tree classification.

W. Gunarathne et. al [9], explained that chronic kidney disease is considered as kidney damage which lasts longer than three months. Identifying CKD in the initial stage is important to provide necessary treatments to prevent or cure the disease. These models have applied on recently collected CKD dataset downloaded from the UCI repository with 400 data records and 25 attributes. Results of different models are compared. From the comparison it has been observed that the model with Multiclass Decision forest algorithm performed best with an accuracy of 99.1% for the reduced dataset with the 14 attributes.

S. Drall, G. S. Drall et. al [10], proposed a machine learning methodology for diagnosing CKD. Chronic kidney disease also recognized as Chronic Renal Disease, is an uncharacteristic functioning of kidney or a failure of renal function expanding over a period of months or years. Habitually, chronic kidney disease is detected during the screening of people who are known to be in threat by kidney problems, such as those with high blood pressure or diabetes and those with a blood relative Chronic Kidney Disease(CKD) patients. So the early prediction is necessary in combating the disease and to provide good treatment.

Problem Statement Definition

Chronic kidney disease (CKD) is one of the most critical health problems due to its increasing prevalence. Chronic kidney disease, also known as chronic renal disease or CKD, is a condition characterized by a gradual loss of kidney function over time. It includes conditions that damage your kidneys and decrease their ability to keep you healthy by filtering wastes from your blood. Diabetes and high blood pressure, or hypertension, are responsible for two-thirds of chronic kidney disease cases. Anyone can get chronic kidney

disease at any age. However, some people are more likely than others to develop kidney disease. Most people may not have any severe symptoms until their kidney disease is advanced.

A better testing method which could possibly detect CKD in the early stages would be much more useful. Medical test results taken for other purposes are used to detect CKD at early stages. Various efforts have been undertaken to advance early therapy to prevent the condition from progressing to chronic disease. Recent research suggests that some of the negative outcomes can be avoided with early identification and treatment. Peculiar and contributing attributes from the above-mentioned test results are combined to develop a Machine Learning Model. This Machine Learning Model will be used to predict CKDs rather early than the presently existing methods.

Ideation & Brainstorming:-

Brainstorming is an activity that will help you generate more innovative ideas. It's one of many methods of ideation—the process of coming up with new ideas—and it's core to the design thinking process.

Brainstorming refers to a problem-solving technique used by teams or individuals. In this process, participants generate various ideas or solutions, then begin discussing and narrowing them down to the best options.

Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that ideation is commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity.

Ideation is the process where you generate ideas and solutions through sessions such as Sketching, Prototyping, Brainstorming, Brainwriting, Worst Possible Idea, and a wealth of other ideation techniques.

Proposed Solution:-

The purpose of this tool is to provide a structured process for identifying a problem, understanding the root causes, ascertaining solution steps, and progress monitoring. With a solution template, you can organize development content that you want to reuse for

customer-specific solutions.

Solution templates enable you to easily start the development of customer-specific solutions, for example, for a specific industry.

The term business model refers to a company's plan for making a profit. It identifies the products or services the business plans to sell, its identified target market, and any anticipated expenses. Business models are important for both new and established businesses.

Problem Solution Fit:-

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem.

Problem-Solution Fit - this occurs when you have evidence that customers care about certain jobs, pains, and gains.

At this stage you've proved the existence of a problem and have designed a value proposition that addresses your customers' jobs, pains and gains.

ADVANTAGES & DISADVANTAGES

Advantages:

Chronic kidney disease (CKD) is one of the most critical health problems due to its increasing prevalence. It is also known as chronic renal disease which is a condition characterized by a gradual loss of kidney function over time. A better testing method which could possibly detect CKD in the early stages would be much more useful using machine learning algorithm

- Greater cost reduction in hospitals for testing
- Helps in early diagnosis of the disease
- Chances of recovery is higher

Disadvantages:

Even Though the CKD prediction model web application consists of a lot of advantages but it comes with certain disadvantages here are some of them .

- Chances of prediction to be wrong for least number of time which can cause problems.
- Vast feature in dataset on discovery of time for the disease making the model.
- Since its a web application it requires scaling of web application.

CONCLUSION

Chronic Kidney Disease as the name suggests it's a chronic disease,any chronic disease would make the person miserable and last longer till their livelihood . If in such cases the disease gets unnoticed in early stages which can be cured by medical facilities it's a huge carelessness and risking a person's life . In such cases finding an optimal solution is important ,thus there comes the use of a machine learningmodel for early detection and prediction of the chronic kidney disease which can greatly reduce the potential risk of getting the disease and get cured immediately if it is detected in early stages of the disease. Think of the traditional way of diagnosing kidney disease,it is through blood test,andblood test reports take longer than expected ,but blood test is not the only step for diagnosing there are still many more tests taken , which can betime consuming . In those cases the model prediction plays an important role in predicting the disease sooner and faster for the medical team to treat the person if he/she is vulnerable.

Thus early detection of chronic kidney disease is very much necessary in current hospital functioning to diagnose the patient in no time and do necessary treatment to cure if found.

FUTURE WORK :

The current work remains the base for the prediction model primarily used by everyone extending from hospitals to normal users . The future aspects can be as follows:

- Scaling the existing application for simultaneous user to request
- Modifying the model based on adding new feature in the existing dataset based on the hospitals input and standards.