International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Chronic Disease Prediction Using Data Sciences

[1] Bhanu Prakash M B
Department of Computer Science and Engineering
VTU CPGS, Mysuru, Karnataka, India

[2]Dr. K THIPPESWAMY, Professor
Department of Computer Science and Engineering
VTU CPGS, Mysuru, Karnataka, India

---***---

Abstract - The prevalence of chronic kidney disease has increased worldwide, making it a major health issue. Predicting chronic kidney disease (CKD) in the current healthcare system is time-consuming due to the increased need for diagnostic, consultation, experience, etc. Early detection of chronic renal disease is a major challenge in modern medicine. Because of renal impairment, harmful substances are not eliminated from the body. Using machine learning algorithms, we primarily strive to detect potentially fatal conditions like chronic renal disease.

Key Words: prediction, healthcare, classification, chronic, retrieval, Naive Bayes.

1. INTRODUCTION

Data Science for the Prediction of Long-Term Illnesses. In the medical field, disease detection is an important topic of study. Predictions of chronic renal “disease cannot be made automatically at this time. Damaged kidneys that are unable to properly filter blood are the hallmark of chronic kidney disease (CKD). Because kidney damage occurs gradually over time, this condition is classified as chronic. This kind of injury can lead to a buildup of toxins in the body. Further medical complications may arise from CKD. Chronic kidney disease (CKD) affects 1o% of the global population and claims the lives of millions every year due to misdiagnosis. The current system is time-consuming and inaccurate because it is performed manually and requires expensive medical equipment. This idea can be put into practise in a healthcare facility, where data from CKD patients can be analysed. It can be used as a platform for virtual health communities, where people can learn about CKD and its many stages. A study group can use it to examine the connection between CKD and its various stages.

2. OBJECTIVES

- The main objective of the proposed system is CKD prediction using machine learning techniques.
- System will generate faster results as we use machine learning techniques.
- System uses supervised learning technique for CKD prediction.
- Systems build as real time application useful for CKD doctors.
- System helps doctors in decision making and to treat the patients in better way.
- System is a generic application meant for different hospitals.

Existing Work Case Study

Benefits provided by the healthcare industry today include insurance fraud detection, patient access to affordable healthcare, the development of innovative treatment methods, efficient hospital resource management, improved patient care and hospital infection control, and a more positive customer experience overall. one of the most important aspects of medicine is the study of disease diagnosis. The healthcare business relies heavily on data mining techniques to analyse enormous amounts of clinical data and make informed judgments. The term "data mining" refers to the method used to discover previously undiscovered data in large datasets. Medical professionals have employed data analysis methods like classification, clustering, regression, and association to detect and predict illness development and guide therapy decisions. Assigning items in a set to predetermined categories is the goal of classification, a supervised learning technique. Data or objects are classified by this procedure into categories whose members share one or more characteristics. SVM, decision tree, Naive Bayes, artificial neural network, etc. are all examples of classification methods. The medical data can be analysed using any of these methods.

3. PROPOSED WORK

Damaged kidneys that are unable to properly filter blood are the hallmark of [3]chronic kidney disease (CKD). Because kidney damage occurs gradually over time, this condition is classified as chronic. This kind of injury can lead to a buildup of toxins in the body. Further medical complications may arise from CKD. Chronic kidney disease (CKD) affects 1o% of the global population and claims the lives of millions every year due to misdiagnosis. Quicker conclusions can be reached with the help of this system. The CKD prediction technique is fully automated. As a practical, real-world solution, the technology has broad applicability throughout healthcare facilities. Predicting chronic kidney disease is possible with the help of supervised learning and the Naive Bayes algorithm/classifier. Useful for clinicians in making more accurate diagnoses of chronic renal disease more quickly".

Scope

- Proposed system is a medical sector application.
- Proposed system area of concern is CKD (chronic kidney disease).
- Proposed system is automation for CKD prediction.
- Proposed system is implemented using .NET technology.
- Proposed system is a real time application.
- System is browser based application accessed from different users and location

4. IMPLEMENTATION AND WORKING

The healthcare business collects vast amounts of raw data in an effort to mine it for useful insights that can improve patient care, diagnosis, and decision making. The severe consequences of CKD make early diagnosis crucial. Chronic kidney disease (CKD) is a major health problem around the world, and its detection has been the subject of numerous studies. The prevalence of chronic kidney disease (CKD) is a major public health issue around the world. Because of renal impairment, harmful substances are not eliminated from the body. our primary goal is developing [1]Classification algorithms for the early detection of serious diseases like Chronic Kidney Disease (CKD). The proposed system would automate the use of the nave bayes classification method for the prediction of chronic renal disease.

Users of the project

Doctor

A physician is someone who defines the parameters for predicting chronic renal disease. Dr. is a client who is receiving care. The system's primary function is to provide a "chronic renal disease prediction" service by analysing patient records.

Receptionist

The details of the patient is registered and entered into the application by the receptionist. This data is available to the doctor through the Doctor login.

Modules of the project

- **Login Module** - here doctor/receptionist gets login to the application by inputting id and password.
- **Upload Patient Data (New patient)** - here either doctor/receptionist inputs the new patient parameters required for CKD prediction.
- **Chronic Kidney Disease Prediction Module** [New patient – Naive Bayes Algorithm] - this the core module where system predicts the CKD for the new patients based on the inputted parameters. Here system uses [2]"Naive Bayes" algorithm for CKD prediction

3-tier Architecture:

3-tier architecture consists of three layers. They are:

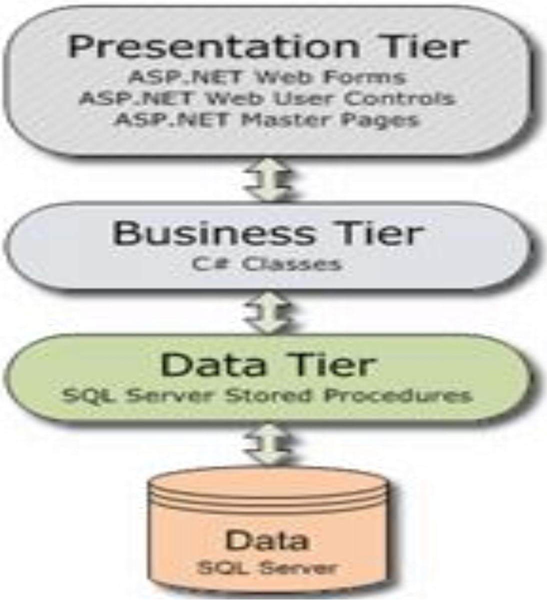


Fig. 1: Three tier Architecture

A. The Data Layer:

In most cases, [2]data is the most important part of an application. The display layer must be fed the data. In a.NET solution, the data layer is a standalone component whose only job is to retrieve data from the database and provide it to the calling code. Logical data reuse is achieved by having many parts of an application call a single data layer method rather than repeatedly implementing the same query. In most cases, this is the most practical option.

B. Business Layer:

Even though a website could bypass the business layer and communicate directly with the data access layer, in practise they rarely do. The business layer is crucial because it checks the validity of the parameters passed to the data layer's methods. This checks the validity of the data before processing it further, and it often verifies the outputs as well. Business rules refer to the guidelines used by the business layer to make "judgments" about the data and thereby validate input. Starting out small and expanding as needed is one of the ideal cases for reusing logic. The business layer aids in the centralization of logic for "maximum reusability."

C. Presentation Layer:

The presentation layer is your ASP.NET website or Windows forms application (the project's user interface). Since the presentation layer is the one that end users interact with directly, it must be considered the most crucial. If the presentation layer is poorly designed, it doesn't matter how effectively the business and data layers are organised; the users will still have a negative impression of the system.

The presentation tier houses the site's UI (User Interface) elements, as well as all the logic that manages the visitor's and client's business interaction. (ASP.NET Web User Controls, ASP.NET Master Pages, and ASP.NET Web Forms)

The display layer sends requests to the business layer, which then responds with an appropriate action based on the business logic it was given. (Classes in C#)

The application's data is stored in the data layer, and it is sent to the business tier when needed. SQL Server Stored Procedures.

High level design:

D. Data Flow Diagram:

The "flow" of data through an IT infrastructure can be visually represented with a data flow diagram (DFD). Data flow diagrams (DFDs) can also be used to depict data processing (structured design). From either an external data source or an internal data store, data items are transferred by an internal process to either an external data sink or another internal data store on a DFD. When looking at a DFD, you won't learn anything about how long each process will take or if they will run in parallel or sequentially. Thus, it differs greatly from a flowchart, which illustrates the direction of control within an algorithm and tells the reader what operations will be carried out and under what conditions, but not what data will be input to and output from the system, nor where the data will be stored (all of which are shown on a DFD).

Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

E. Symbols used in DFD's:

Data Flows:

A data flow is the chain of events that leads from one thing to another. It stands in for the values of data that are used as placeholders in a computation. Between the process and the recipient of the data value, an arrow is drawn. The name or nature of the information being conveyed is typically labelled along the arrow.

Processes:

A process transforms data values. The lowest processes are our functions without side effects

Actors:

An actor is a value producer or consumer that actively contributes to the data flow graph. In a dataflow graph, inputs and outputs are connected to actors. Actors are frequently referred to as terminators since they are located on the edge of the flow graph but end the flow of data as sources or sinks.

Data Store:

A data store is a node in a data flow diagram that acts as a repository for information. In contrast to an actor, a data store only acts in response to requests for storing and retrieving information.

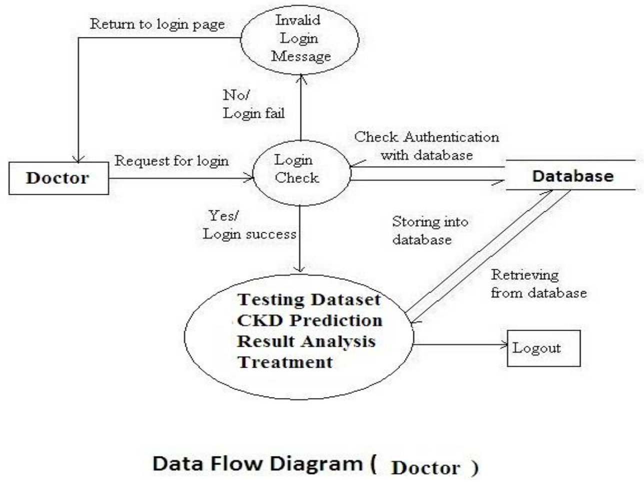


Fig. 2: Data flow diagram

F. Naïve Bayes Algorithm Steps

Step 1: Scan the dataset (storage servers)

retrieval of required data for mining from the servers such as database, cloud, excel sheet etc.

Step 2: Calculate the probability of each attribute value. [n, n_c, m, p]

Here, we apply the following formula to each property to determine its likelihood of recurrence. (to be discussed further on). We need to use the formulas for each category of illness.

Step 3: Apply the formulae

P(attributevalue(ai)/subjectvaluevj)=(n_c + mp)/(n+m)

Where:

- n = the number of training examples for which v = vj
- nc = number of examples for which v = vj and a = ai
- p = a priori estimate for P(aijvj)
- m = the equivalent sample size

Step 4: Multiply the probabilities by p for each class, here we multiple the results of each attribute with p and final results are used for classification.

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

G. Sample Example

Attributes (Parameters) – AGE, GENDER, BP [m=3]
Subject (Disease) – CKD, NoT CKD [p=1/2=o.5]

H. Training Dataset

Patient Name	Age	Gender	Blood Pressure	Random blood sugar	Serum Creatinine	Disease (subject)
Akash	25	M	110	281	2.4	CKD
Aruna	25	F	112	192	1.9	CKD
Shilpa	30	F	120	163	1.1	NoT CKD
Kumar	35	M	105	336	1.6	CKD
Chaitra	35	F	120	135	1.0	NoT CKD

Fig. 3: Training Dataset

New Patient data – Rajani Parameter (AGE-25, GENDER - Female, BP-112)

Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Disease – CKD /NOT CKD

P=[n_c + (m*p)]/(n+m)

CKD	NoT CKD
25 P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.7	25 P=[n_c + (m*p)]/(n+m) n=2, n_c=0,m=3,p=0.5 p=[0+(3*0.5)]/(2+3) p=0.3
Female P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.7	Female P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.3
112 P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5	112 P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5

Fig. 4: Calculation steps

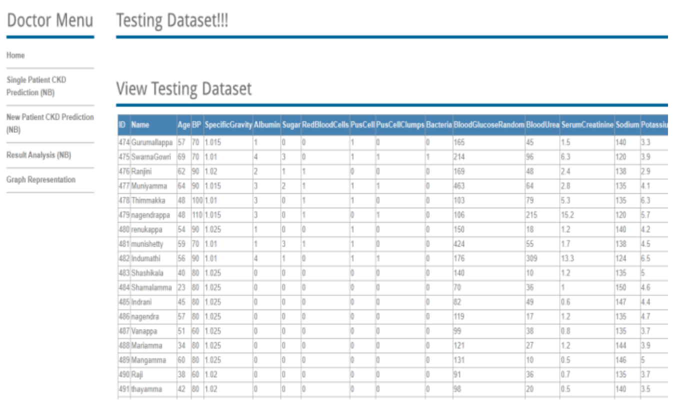
CKD = o.7 * o.7 * o.5 * o.5 (p) = o.1225

NoT CKD = o.3 * o.3 * o.5 * o.5 (p) = o.o225

Since CKD > NoT CKD

So this new patient is classified to CKD

I. Testing Dataset



5. CONCLUSION

This project is a healthcare industry application that aids doctors in predicting CKD disorders using CKD indicators. Using automation for CKD disease prediction, the illness and its subtypes and consequences can be quickly and cheaply identified from the clinical database. The Nave Bayes classification technique is used to get the job done. The field of data mining includes this method of classification. This system utilises data from chronic kidney disease patients over time to make disease predictions.

The programme can be tweaked to forecast more chronic diseases using different variables. Chronic ailments like heart disease, cancer, diabetes, stroke, obesity, and arthritis are all within the scope of this app's predictive abilities. The functionality of the web app can be improved so that it can be used on Android smartphones.

REFERENCES

1. H. Witten and E. Frank, "Data Mining Practical Machine Learning Tools and Techniques," 2nd ed., San Francisco/ABD, 2oo5
2. Bruce D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli and I. o. Ellis, "A 'non-parametric' version of the naïve Bayes classifier," Knowledge- Based Systems, vol. 24, 2o11, pp. 775-784.
3. https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease (Access Date: 2o18 February 7).
4. C. Cortes and V. Vapnik, "Support- Vector Networks," Machine Learning, vol. 2o, 1995, pp. 273-297
5. http://www.datascience.istanbul/2o17/o7/o2/hata-matrisini-confusionmatrix-yorumlama (Access Date: 2o18 February 7)
6. Naïve bayes model for analysis of voting rate failure in election system, K Gururaj, K Thippeswamy - International Journal on Recent and Innovation

.