Kidney Disease Prediction Using Data Mining Classification Techniquesand ANN

Sahana B J, Dr Minavathi

M.Tech Student, Dept. of C.S., P.E.S College of Engineering, Mandya, India

HOD, Dept. of I.S., P.E.S College of Engineering, Mandya, India

ABSTRACT: Information mining is a procedure of separating helpful data from gigantic measure of dataset. Information mining has been a present pattern for getting analytic outcomes. In therapeutic application, tremendous measure of unmined information is gathered by the social insurance industry with a specific end goal to find concealed data for successful conclusion and basic leadership. There are numerous information mining systems like grouping, clustering and so on. The goal of our paper is to anticipate Chronic Kidney Disease (CKD) utilizing arrangement methods like Naive Bayes and to foresee the phases of endless kidney illness utilizing the Artificial Neural Network (ANN) like C4.5.

KEYWORDS: Data mining, Classification, Chronic Kidney disease, Naive Bayes, Artificial Neural Network,

I. INTRODUCTION

Information mining is a procedure of separating valuable data from enormous measure of dataset. Information mining has been in numerous areas like picture mining, assessment mining, web mining, content mining, diagram mining and so on. Furthermore it has turned out to be prevalent in wellbeing association as there is a prerequisite of expository strategy for foreseeing and discovering obscure examples.

Information mining assumes an imperative part in finding new patterns in social insurance industry. It is especially valuable in wellbeing field when no accessibility of affirmation supporting a specific treatment choice is found. It contains an arrangement of devices and procedures which when connected to prepared information, gives valuable data to medicinal services experts for settling on right choices and that enhances the execution of patient administration errands. Patients with same medical issue can be consolidated together and good treatment arrangements could be given in view of information gathered from past patient.

The worldwide medical issue which has been confronted now days is constant kidney infection (CKD) this is the region of concern. Interminable kidney illness is a condition where kidneys end up plainly harmed and can't channel poisonous materials in our body. Our work dominatingly concentrates on identifying life debilitating ailments like Chronic Kidney Disease (CKD) utilizing Classification calculations like Naive Bayes and Artificial Neural Network(ANN) like C4.5 to predicts phases of Chronic kidney disease(CKD).

II. RELATED WORK

- In Previousmethods[5]Chronic Kidney disease with unknown aetiology (CKDu), were critically reviewed and harnessed with other relevant literature with a view to justify the possible causative reason(s) for this chronic health problem. To this end, authors first reviewed the literature on socio-economic, environmental, meteorological, health and geo-environmental/chemical aspects in relation to the CKD of unknown aetiology in Sri Lanka and then studied the known a etiology of CKD (key medical facts) in detail highlighting a comprehensive list of causes.

- Based on the nature of economic development in the CKDu prevailing area, previous studies have suggested that the causative factor for CKDu is probably environmental and is related to the nature of anthropogenic activities. Therefore, authors investigated subject more into the chronic exposure to heavy metals associated

with agricultural activities. Potential sources of heavy metals in the area are mentioned. The geochemical mobility and health effects of three selected elements namely arsenic (As), cadmium (Cd) and Lead (Pb) are studied in detail.

- The current scientific observations reported on CKD were evaluated with established chemical, geochemical, and health risk of As, Cd and Pb. According to the review, it can be justified that CKDu of the North Central Province (NCP) in Sri Lanka is a decease caused by the chronic exposure and cumulative effects of elevated levels of heavy metals associated with agricultural activities.

III. PROPOSED SYSTEM

The figure. 1., contains four actors: Admin, receptionist, doctor and patient and also their modules, database, server. The proposed system is a tool for healthcare industries. Admin plays a major role in creating new doctor and receptionist, manages the staffs , adds the attributes, adds the values to the attributes, adds the stages, adds the disease type.
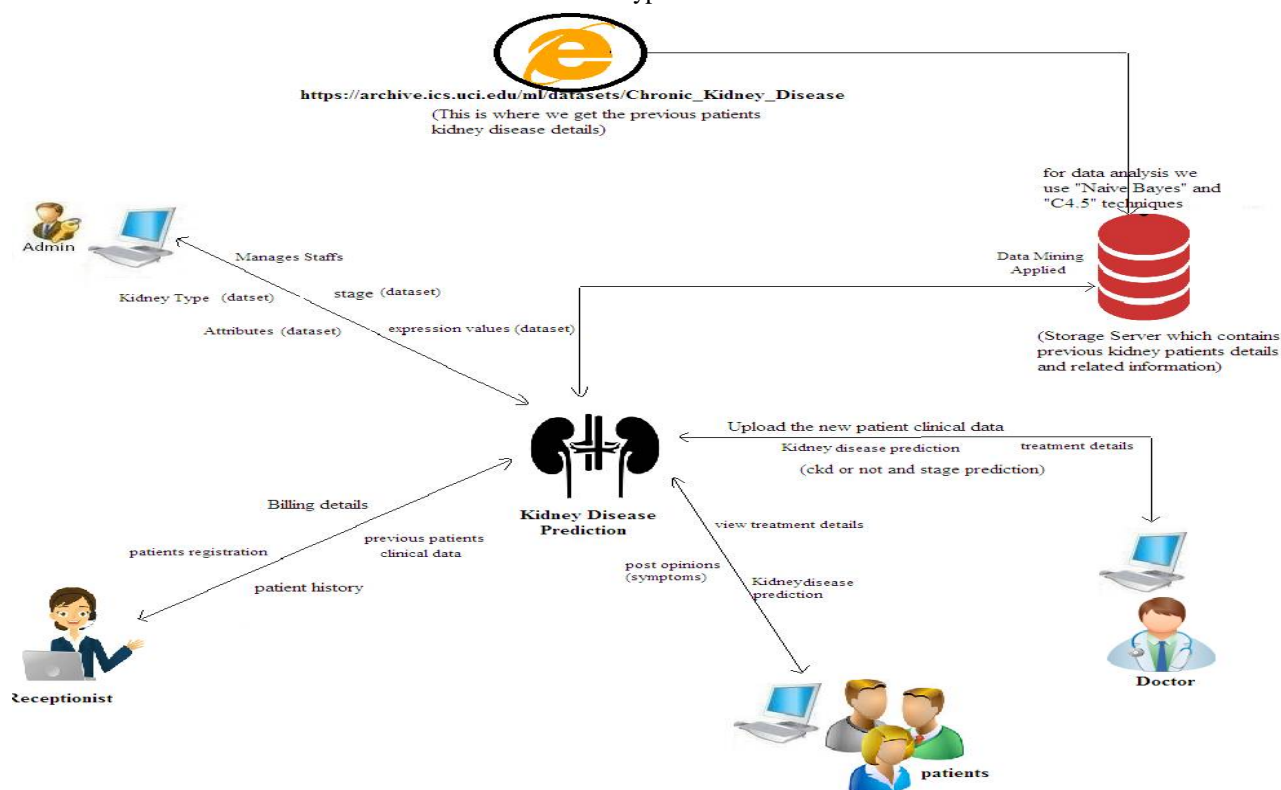

Figure 1:Architecture of Proposed system

Receptionist can login with the given userid and password and registers the patient by providing userid and password, he can report the details of who all suffering from ckd and who are not having ckd.

Doctor can login with valid userid and password and register patient, adds old patient history, reporting can done, and can provide the treatment to the patients.

Patient can login with userid and password and can view his details and disease he is suffering from and also the treatment details.

Firstly the admin adds the doctor and receptionist and adds the attributes and its values on which the ckd depends on later the doctor adds the old patient history with the previous details of the patient.

In order to predict the disease for new patients first we need to add the new patient and add the attribute values of new patient and then click on predict (naïve) then we get result as CKD or NOTCKD. If the patient is CKD then we get other button predict (c4.5) these is to predict the stages of ckd that is from STAGE1 to STAGE5.

Here in this project the new patient disease prediction depends on the old patients entered and todays new patient becomes the old patient for tomorrow the dynamic table will be generated.

All the actors have the privilegeto change the old password.

PROPOSED ALGORITMS

A. *Methodology for Disease prediction:*

Proposed framework makes use of "Naïve Bayes Algorithm" This algorithm predicts whether the patient is suffering from ckd or notckd.

Step 1: Scan the dataset (storage servers)

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

Step 3: Apply the formulae

$$P(attributevalue(a_i)/subjectvaluev_j)=(n\_c + mp)/(n+m)$$

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

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

B. *Methodology for Stage Prediction:*

Proposed framework makes use of "C4.5" algorithm. This algorithm predicts the stages of ckd.

Step 1: Scan the dataset (storage servers)

Step 2: for each attribute a, calculate the gain [number of occurrences]

Step 3: Let a_best be the attribute of highest gain [highest count]

Step 4: Create a decision node based on a_best – retrieval of nodes[patient] where the attribute values matches with a_best.

Step 5: recur on the sub-lists [list of patient] and calculate the count of outcomes[Stages] – termed as subnodes. Based on the highest count we classify the new node.

IV. RESULT

Here first we need to adds the old patient history figure2 shows adding the attributes values. Later by using these attributes of the old patients i.e., figure 3 that contain the old patient entries along with these past clinical details we predict whether the new patient is suffering from ckd or not in figure4 . If he is suffering from ckd which stage he is suffering this all can be done in this project.
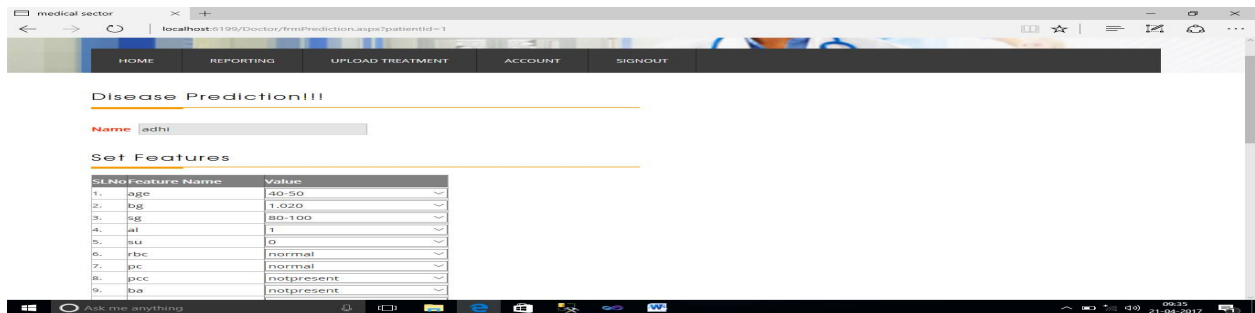
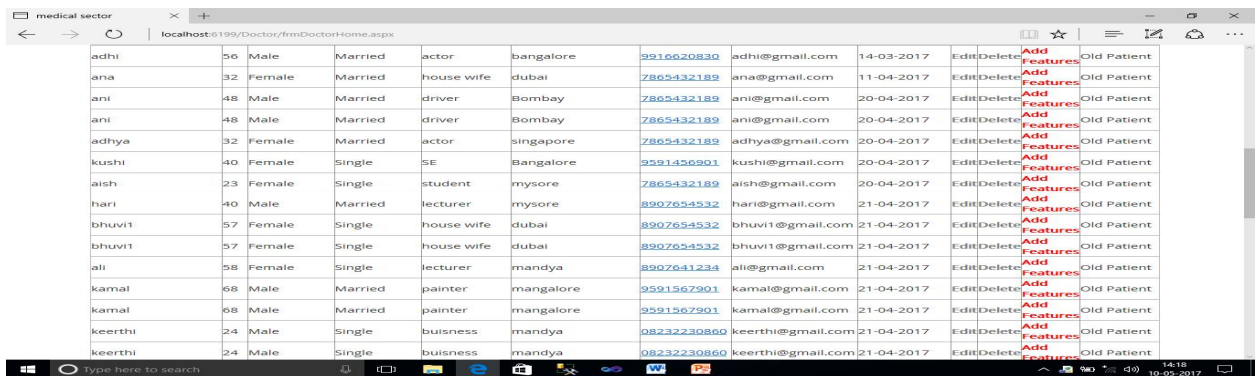Figure. 2. Adding attribute values to the attributes
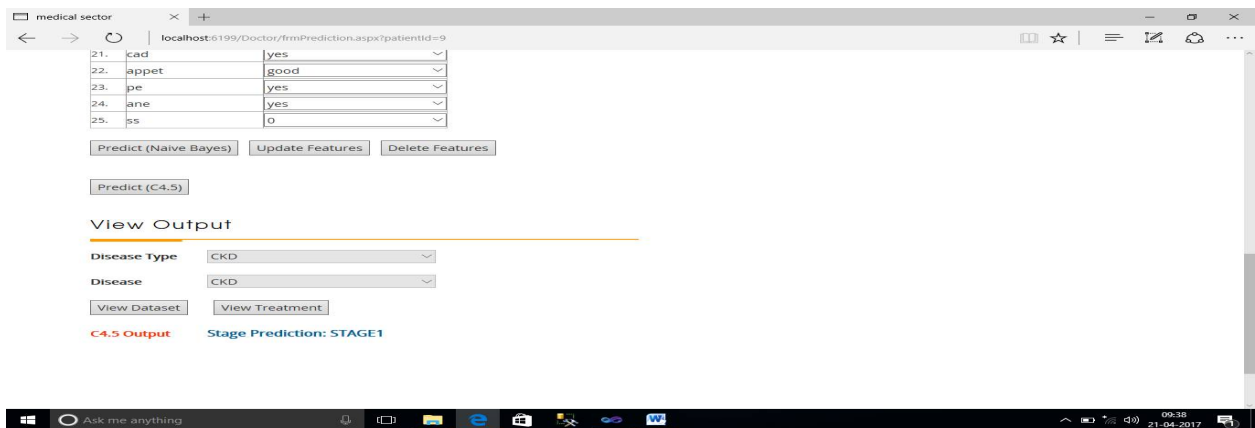

Figure 3: past clinical data of old patients

Figure. 4 Prediction of CKD along with its stage

V. CONCLUSION AND FUTURE WORK

This project is a medical sector application which helps the medical practitioners in predicting the disease types based on the symptoms. Patients can also predict diseases by entering symptoms in the form of sentences. It is automation for disease prediction and it identifies the disease, its types and complications from the clinical database in an efficient and an economically faster manner.It is successfully accomplished by applying the Naïve Bayes algorithm for classification and C4.5 for stages prediction. This classification technique comes under data mining technology. This algorithm takes symptoms as input and predicts the disease based on old patients data.

SMS/Email Module – In the proposed system, admin assigns Id and password for doctors and receptionists and is intimated manually, so we can add SMS/Email module as a future enhancement where doctors and receptionists receives an SMS or Email regarding the Id and password.

Query Module- we can add the query module as a future enhancement to the application where doctor, receptionist and admin of the application can interact with each other.

REFERENCES

[1] Aditya Sundar1, P. Pushpa Latha2, M. Rama Chandra3[et.al, 2012],Performance Analysis Of Classification Data Mining Techniques Over Heart Disease Data Base.

[2] An empirical study of the naive Bayes classifier by I. Rish T.J. ,Watson Research Center
Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[3] Harleen Kaur , Siri Krishan Wasan and Vasudha Bhatnagar, THE IMPACT OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSTICS, Data Science Journal, Volume 5, 19 October 2006 pp119-126.

[4] Data Mining Concepts, backgrounds and methods of integrating uncertainty in data mining, Yihao Li, Southeastern Louisiana University

[5] Sunethra Kanthi Gunatilake1*, S. Sunil Samaratunga2 and Ruvini Takshala Rubasinghe1,Chronic Kidney Disease (CKD) in Sri Lanka - Current Research Evidence Justification: A Review,Sabaragamuwa University Journal Volume 13 Number 2; December 2014.

BIOGRAPHY

Sahana B Jis a final year student of Master of Technology (M.Tech) Pursuing in Computer Engineering, P.E.S college of Engineering, Mandya, Karnataka, India. She received Bachelor of Engineering (BE) from Maharaja Institute of Technology, Mysore, Karnataka, India. Her research interests are Big Data and Mining Tools, Algorithms, Data Analytics, web 2.0 etc.

Dr Minavathi is anHead of the Dept. in the Information Science Department, P.E.S College of Engineering, Mandya. She received bachelor degree from P.E.S College of Engineering, Mandya and She received Master of Technology (MS) degree fromBITS, PILANI, India. She got PhD from Mysore University. Her interest in Medical Image Processing.