A Novel Algorithm for Prediction of Chronic Kidney Risks using Machine Learning Schemes

Sreeji, Balamurugan Balusamy

Abstract: The goal is to assess the initial prediction of severe kidney ailments termed as severe renal ailments for diabetic patients with the aid of machine learning schemes and lastly recommending choice hierarchies to derive solid outcomes with predictable precision by estimating their performance based on its requirements and sensitivity. The behaviour of learning scheme is based on a collection of information extracting pointer with balanced outcomes on the created model. Identifying the knowledge from elaborated repositories termed as information extraction. Apart from learning the prevailing clinical base heart ailment information set 600 clinical records are gathered from diabetic analysis centre globally. The information set is verified for categorization based on Naive Bayes and choice based Hierarchies. Based on evaluating the categorization scheme in terms of Naive Bayes and choice based hierarchies it is possible to arrive at a conclusion with precision of 90.2% for choice based hierarchical categorization. For improving the precision of the estimation outcomes it uses schemes like neural network and grouping information that highly aids for meeting our goals and offers chances for forthcoming analyses.

Keywords – Categorization, Machine Learning, Naive Bayes, Choice Based Hierarchies and Chronic Ailments.

I. INTRODUCTION

The information extraction is gaining its attention these days in the domain of healthcare, deception and misuse detection. Categorization is an attractive information extraction function for addressing the items as a set to focus classes or types [1]. The kidney failure remains one amongst the various types like heart disease and blindness which are the outcomes of chronic ailments. Dialysis is the only possible way to uphold the functioning of kidneys unnaturally and it is also a throbbing and money involving process [3] [4] [5]. According to the World Health Organization more than millions of individuals around the globe are prone to serious kidney-related issues that are getting increased every year. Hence, the timely prediction is needed without delays so that necessary cautions or controls can be made on time [6].

To acquire the much-needed data from the medical repositories information extraction is very handy. By aggregating the machine learning and analytical strategies smartly needful data can be extracted from the medical repositories.

II. TECHNIQUES

The information extractions schemes were previously employed by various scholars for identifying the ailments. The detection of these various ailments is performed in combination by making using and assessing machine learning, statistical and medical repository practice [7] [8].

Machine learning schemes can work in harmony with several of the analyses and repositories which in turn aids in mining the concealed patterns and associations from immense and various differential information. For assuring the precision of the selected classifier the prevailing tests can be authenticated. Additionally, these elements such as precision, sensitivity and definiteness are usual for identification of ailments. Based on the Naive Bayes and Choice-based hierarchies for the intended categorization schemes it is possible to accomplish the outcomes of detecting the kidney related problems at the beginning phase.

Fig 1. Proposed Prototype

There prevail two varied schemes namely categorization and assessment comprised under information extraction scheme aiding to construct information for instructions, categorized analytical prototype, and validating the effectiveness of categorization [5]. The assistance for forthcoming activities can also be obtained and acquired from the aforementioned categorized information [9] [10].
Various sectors also make use of rapid extraction software for their purpose and validation resulting in effective authority. The rapid extraction software is also employed in validation and effortless prototyping that aids in synchronizing the actions of machine learning, information extraction, text extraction, statistical analyses, and business analytics [12]. These two information extraction tools namely WEKA and YALE make use of Java providing an abundance of opportunities for the field of business, knowledgebase, scientific and medical research. The rapid extraction software is greatly helpful in modeling 10 cross-validations of a machine learning scheme for improving the performance level. The validation is carried out by splitting the information into 10 segments comprising 90% of the actual information. The cross verification is of great help for developing a set of information folds for instructions. The evaluations are carried out with an anticipated rate of errors over the outcomes by making use of the data employed for instructions. The element-wise information of each record is depicted in table 1.

Table 1: Elements Used for Assessment

| Risk Factor | Elements     | Variety | Details               |
|-------------|--------------|---------|-----------------------|
| Gender      | Gender       | Binomial| Male/Female           |
| Age         | Age          | Integer | Patient’s Age         |
| Genetics    | Genetics     | Polynomial| Father, Mother or Both|
| Weight      | Weight       | Number  | Patient’s Weight      |
| Smoking     | Smoking      | Number  | Yes/No                |
| BP          | BP           | Polynomial| Patient’s Blood Pressure|
| Fasting     | Fasting     | Integer | Fasting Blood Sugar   |
| PP          | PP           | Integer | Post Prondial Blood Glucose |
| AIC         | AIC          | Number  | Glycosylated Haemoglobin Test |
| LDL         | LDL          | Integer | Low-Density Lipoprotein |
| VLDL        | VLDL         | Integer | Very Low-Density Lipoprotein |
| HDL         | HDL          | Number  | High-Density Lipoprotein |
| Urea        | Urea         | Number  | Urea Creatine         |
| Threat Category | Threat Category | Polynomial| Threats of Kidney Disease |

III. NAVIE BAYES SCHEME

The hypothesis is employed based on the Naive Bayes categorizer based on the supposition to address the conditional and easy possibilities. Based on the Bayes hypothesis the possibility of the event happened is located. In case if ‘X’ represents the earlier event and ‘Y’ represents the reliant event, then the hypothesis is stated as,

\[ P(Y \text{ given } X) = \frac{P(X \text{ and } Y)}{P(X)} \]  

To estimate the prototype needed for categorizing and deciding the difference of variables for each category lesser items of instruction information are needed as the reliant variables are presumed. The Naive Bayes task is depicted in fig. 2.

IV. NAIVE BAYES PERFORMANCE

V. CHOICE BASED HIERARCHIES

It is an assumption scheme employed to build a categorization or regression prototype in a hierarchical model. It breaks into much smaller subsets and also parallel creates choice nodes and leaf nodes. The two types of choice-based hierarchies employed for information extraction are assessing categorized hierarchies and assessing regression hierarchies for various expected results like fitting to a precise information class or real number. The analytical categorizer C4.5 scheme is designed to generate a choice based hierarchies as portrayed in fig. 3.

VI. RESULT AND DISCUSSIONS

A prototype with a machine learning scheme along with validation is generated with great precision. The prepared subset is validated out of a categorizer is created. Therefore, the ratio of categorized information is obtained. With the aid of Naive Bayes categorization scheme usual scattering possibilities are presumed. The fundamental elements are located for estimating based on the Naive Bayes scheme in terms of age, gender, smoking, alcohol, cholesterol HDL, etc. The threat is marked as high, average and low for the age group of 50 – 54 which are represented over the x and y axes as in fig. 4. Here, the x-axis represents age and the y-axis represents density respectively.
Based on the estimated values of the aforementioned elements the age element are allocated and is portrayed in the form of categorization. In table 2, \( l_t \), \( a_t \), and \( h_t \) represents low, average and high levels of threats respectively.

Table 2: Naive Bayes Scattering Table for Age

| Elements | Metrics   | Low   | Average | High   |
|----------|-----------|-------|---------|--------|
| Age      | Mean      | 52.25 | 51.75   | 50.52  |
| Age      | Standard  | 10.05 | 9.88    | 9.56   |
|          | Deviation |       |         |        |

With the aid of true positive \( (t_p) \) and true negative \( (t_n) \) categorizations, the performance of the prototype is estimated. For multi-category categorization issues, it is possible to portray true positive, false positive, true negative and false negative for every category of ‘c’. Identically it portrays the calculations \( T_{pc}, F_{pc}, T_{nc} \) and \( F_{nc} \) for each category of ‘c’. The precise metrics can be estimated for evaluating the multi-category categorization outcomes respectively. The rate of true positive, accuracy and \( f \) – estimate values for each category and precision are estimated.

\[
Rate_T = \frac{T_p}{T_p + F_n}
\] (2)

The rate is the metric for estimating the way these categories could be estimated with the true members for precise categories. The rate of true positive is not adequate for comprehensively estimating the performance of the categories in a single category and hence it is possible to estimate the accuracy of class ‘c’ as,

\[
Accuracy(c) = \frac{T_p}{T_p + F_p}
\] (3)

The \( f \) – rank or \( f \) – the calculation is estimated based on the precision of validation and it is the harmonic ways for accuracy and recall that is estimated as,

\[
f = 2 \left( \frac{\text{Accuracy} \times \text{Recall}}{\text{Accuracy} + \text{Recall}} \right)
\] (4)

The \( f \) – calculation can also be estimated for category ‘c’ as,

\[
f = 2 \left( \frac{\text{Accuracy}(c) \times \text{rateof}\{T_p\}}{\text{Accuracy}(c) + \text{rateof}\{T_p\}} \right)
\] (5)

The performance of the Naive Bayes categorizers for the information sets are depicted in table 3. From table 3 the \( t_l, t_a, \) and \( t_h \) represents the true low, true average and true high category values and \( a_l, a_a, \) and \( a_h \) representing accuracy as low, average and high category values. The varying levels of precision in terms of gain rate, data expansion and Gini directory are in the forms of choice based hierarchies based on the diverse division schemes as depicted in table 4.

Table 3: Naive Bayes Scattering Accuracy

| Divide Scheme | Precision (%) | Categorization Faults (%) |
|---------------|---------------|--------------------------|
| Gain Percentage | 82.25         | 10.08                    |
| Data Gain     | 91.5          | 9.09                     |
| Gini Directory | 72.25         | 22.35                    |
The choice-based hierarchies are portrayed in table 4 offering the outcomes of the differed information, the precision of which is estimated to almost 90.2%. From table 5, the $t_l$, $t_a$, and $t_h$ represents the true low, true average and true high category values and $a_l$, $a_a$, and $a_h$ representing accuracy as low, average and high category values. Table 6 permits the evaluation of the Naive Bayes scheme, the choice-based hierarchies' scheme employed for the two schemes resemble to be more efficient as the accurate assessments are portrayed as 91.2% for the patients with chronic kidney ailments.

Table 5: Performance of Choice Based Hierarchies with
Precision of 90.2%

| Category | $t_l$ | $t_a$ | $t_h$ | Precision |
|----------|-------|-------|-------|-----------|
| $a_l$    | 625   | 15    | 7     | 90.0      |
| $a_a$    | 10    | 120   | 15    | 91.2      |
| $a_h$    | 2     | 8     | 90    | 91        |
| Recollection | 85 | 82    | 80    | 90.2      |

The overall intention is for the information extraction scheme for various types of assessment of medical related information. The performance of the choice based hierarchies is predicted to be 90.2% precise as estimated against Naive Bayes scheme. The categorization scheme for diabetic information set performance is acquired as. It is also located that the extraction is aid to revive synchronization from elements that are not straightforward representation of the categories that are attempting to assess. It is focused on estimation on improving the performance of assessment system where the precision in neural networks and grouping scheme aid in information assessment.

REFERENCES
1. A Manish Kumar, —Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm‖ International Journal of Computer Science and Mobile Computing, Vol. 5, Issue. 2, February 2016.
2. S.Ramya, Dr. N.Radha, —Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms‖ International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2016.
3. Jyoti Saini, R.C Gangwar and Mohit Marwaha, —A NOVEL DETECTION FOR KIDNEY DISEASE USING IMPROVED SUPPORT VECTOR MACHINE‖ International Journal of Latest Trends in Engineering and Technology Vol.(7)Issue(4), 2015.
4. Parul Sinha, Poonam Sinha, — Comparative Study of Choronic Kidney Disease Prediction using KNN and SVM, International Journal of Engineering Research and Technology, Vol(4), Issue-12, 2015.
5. Neha Sharma, Er.Rohit Kumar Verma, — Prediction of Kidney Disease by using Data Mining Techniques International Journal of Advanced Research in Computer Science and software Engineering, Vol 6, Issue 9, September 2016.
6. Dr. S. Vijayarani, Mr. S. Dhayanand, — KIDNEY DISEASE PREDICTION USING SVM AND ANN ALGORITHMS — International Journal of Computing and Business Research(IJCBR), Volume 6, Issue2, March 2015.
7. Sai Presad Potharaju, M.Sreedevi, — Ensembled Rule Based Classification Algorithms for predicting Imbalanced Kidney Disease Data International Journal of Engineering Science and Technology Review 9(5) (2016).
8. Lambodar Jena, Narendra Ku. Kamila, — Distributed Data Mining Classification Algorithms for Prediction of Chronic-Kidney-Disease — International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-11).
9. L.Jerlin Rubini, Dr.P.Eswaran, —Generating comparative analysis of early stage prediction of Chronic Kidney Disease International Journal Of Modern Engineering Research (IJMER) ISSN: 2249–6645, Vol. 5, Iss. 7, 2015.
10. Morteza Khavání Zadeh, Mohammad Rezapour, Mohammad Mehdi Sepehr, —Data Mining Performance in Identifying the Risk Factors of Early Arteriovenous Fistula Failure in Hemodialysis Patient International Journal of Hospital Research, 2012.
11. Basma Boukenenze, Hajar Mousannif and Abdelkrim Haqiq, — Performance of Data Mining Techniques to Predict in Healthcare Case Study: Chronic Kidney Failure Disease, International Journal of Database Management Systems (IJDMIS), Vol.8, No.3, June 2016.
12. Pushpa M. Patil, —Review on Prediction of Chronic Kidney Disease using Data Mining Techniques, International Journal of Computer Science and Mobile Computing, Vol. 5, ISSN 2320–088X, Issue. 5, May 2016.
13. Mohammad Rezapour, Morteza Khavání and Mohammed Mehdi Sepehr, —Implementation of Predictive Data Mining Techniques for Identifying Risk Factors of Early AVF Failure in Hemodialysis Patients Computational and Mathematical Methods in Medicine, Volume 3, 2013.
14. R.Sujatha, Dr. Ezhilmaran, —PERFORMANCE ANALYSIS OF DATA MINING CLASSIFICATION TECHNIQUES FOR CHRONIC KIDNEY DISEASES International Journal Of Pharmacy & Technology, ISSN: 0975-766X, Vol.-6, 2016.
15. F Al-Turjman, MH Nawaz, UD Ulusar, Intelligence in the Internet of Medical Things era: A systematic review of current and future trends ,Computer Communications 2019
16. M Bhouaneswari, GN Balaji, F Al-Turjman, Machine Learning Parameter Estimation in a Smart-City Paradigm for the Medical Field,Smart Cities Performability, Cognition, & Security, 139-151
17. BD Deebak, F Al-Turjman, M Aloqaily, O Alfandi, An Authentic-Based Privacy Preservation Protocol for Smart e-Healthcare Systems in IoT, IEEE Access 7, 135632-135649

AUTHORS PROFILE

Sreeji.S received the M.E degree in Computer Science and engineering from Satyabama University, Chennai, India, in 2014 and the B.E degree in Computer Science and engineering from Anna University, Chennai, India, in 2009. He is currently pursuing the Ph.D. degree in Computer Science and engineering at Galgotias University, G.Noida, INDIA. From 2009 to 2012, he was a Software Engineer, Tenon Services, Bangalore. From 2014 he was working in various university. Currently he was working in Galgotias University G.Noida, INDIA.

Dr. Balmurugan Balusamy received the Ph.D from VIT University. He had published more than 90 papers as part of my research work and research interest. Currently he is working in Galgotias University Greater Noida, INDIA.