Applications of Data Mining Techniques in Prediction of Heart Attacks Using Naïve Bayes and Rule Based Classification Algorithm

S.Kavitha[1] G.Ramya[2]
[1,2]Research Scholar

Abstract— Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, we briefly examine the potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to massive volume of healthcare data. This is an extension of naïve Bayes to imprecise probabilities that aims at delivering robust classifications also when dealing with small or incomplete data sets. Discovery of hidden patterns and relationships often goes unexploited. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease.

Key words: Rule based, Decision tree, Naïve Bayes, Artificial Neural Network, data mining

I. INTRODUCTION

Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases is given as follows: "Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data". Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. The discovered knowledge can also be used by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives.

II. PROBLEM DEFINITION

Cardiovascular disease (CVD) causes serious illness, disability, and death. Over 70 million Americans have some form of cardiovascular disease. Heart disease and stroke combined are responsible for 40% of the deaths in the nationeach year and are the first and third leadingcauses of death, respectively. the parameters of heart attack prediction with corresponding values and their weightages. In that, lesser value (0.1) of weightage comprises the normal level of prediction and higher values other than 0.1 comprise the higher risk levels. If the Male And age < 30 And Smoking = Never And Overweight = No And Alcohol = Never And Stress = No And High saturated fat diet (hsfd) = No And High salt diet (hsd) = No And Exercise = Normal And Sedentary Lifestyle (Inactivity) = No And Hereditary = No And Bad Cholesterol = Low And Blood Sugar = Normal And Blood Pressure = Normal And Heart Rate = Normal 650 Shantakumar B.Patil and Y.S.Kumaraswamy Or Male And age > 50 and age < 70 And Smoking = Current And Overweight = No And Alcohol = Past And Stress = No And High saturated fat diet (hsfd) = No And High salt diet (hsd) =Yes And Exercise = High And Sedentary Lifestyle (Inactivity) = No And Hereditary = No And Bad Cholesterol = Low And Blood Sugar = Normal And Blood Pressure = Normal And Heart Rate = Normal Then Risk Level = Normal Otherwise If Male And Age > 30 and age < 50 And Smoking = Current And Overweight = Yes And Alcohol = Current And Stress = Yes And High saturated fat diet (hsfd) = No And High salt diet (hsd) = Yes And Exercise = High And Sedentary Lifestyle (Inactivity) = Yes And Hereditary = Yes And Bad Cholesterol = High And Blood Sugar = High And Blood Pressure = Low And Heart Rate = Low Or High.

III. METHODALOGY

A. *Naive-bayes classification algorithm*

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data.

An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

1) *Bayes theorem*

The theorem describing how the conditional probability of each of a set of possible causes for a given observed outcome can be computed from knowledge of the probability of each cause and the conditional probability of the outcome of each cause.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = P(A \cap B)/P(A)$$

$$P(A/B) = P(A \cap B)/P(B)$$

$$P(A \cap B) = P(B/A)/P(A) = P(A/B)\,P(B)$$

From the formulas the Bayes Theorem States the Prior probability: Unconditional probabilities of our hypothesis before we get any data or any NEW evidence Also stated is the posterior probability: A conditional probability about our hypothesis (our state of knowledge) after we revised based on the new data. Likelihood is the conditional probability based on our observation data given that our hypothesis holds.

$$\text{Posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{Evidence}}$$

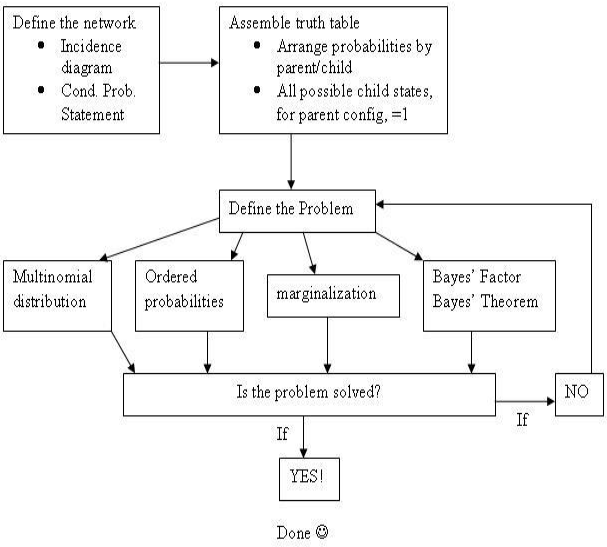


Fig. 1: The Flow of Processes Starting from the Naïve Bayes Classification

2) Pseudo code

Calculate diagnosis="yes", diagnosis="no" probabilities P yes, P no from training input.

For Each Test Input Record

For Each Attribute

Calculate Category of Attribute Based On Categorical Division

Calculate Probabilities Of Diagnosis="Yes",

Diagnosis="No" Corresponds To That Category P(Attr,Yes),

P(Attr,No) From Training Input .

For Each Attribute

CalculateTheResult yes=Result yes*

P(Attr,Yes),Result no= Result no*P(Attr,No);

Calculate Result yes= Result yes *P yes

Resultno= Resultno*P no;

If(Result yes > Result no)Then Diagnosis="Yes";Else Then Diagnosis ="No";

B. Rule-based classification

The basic steps for rule-based classification are as follows. Specific steps are explored in greater detail in the example.

(1) Create a table for the documents to be classified, and populate it.
(2) Create a rule table (also known as a category table). The rule table consists of categories that you name, such as "medicine" or "finance," and the rules that sort documents into those categories.

These rules are actually queries. For example, you might define the "medicine" category as consisting of documents that include the words "hospital," "doctor," or "disease," so you would set up a rule of the form "hospital OR doctor OR disease." Classify the documents.We propose a new rule-based algorithm for classifying and predicting both certain and uncertain data. We integrate the uncertain data model into the rule-based mining algorithm. We propose a new measure called probabilistic information gain for generating rules. We also extend the rule pruning measure for handling data uncertainty. Rule-based classifiers make use of set of IF-THEN rules for classification. We can express the rule in the following from, IF condition THEN conclusion

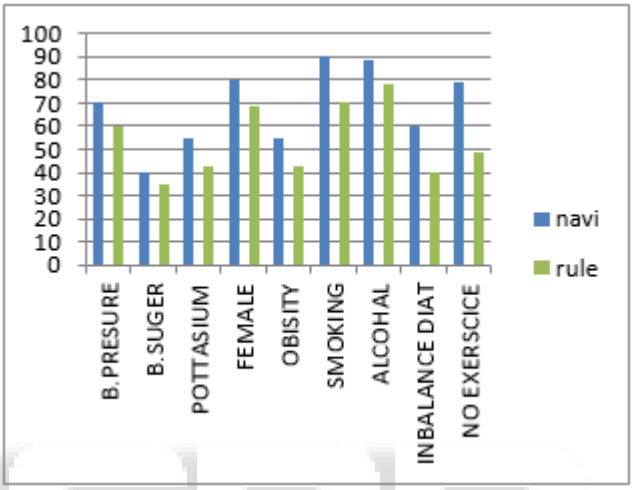
Let us consider a rule R1,

We can also write rule R1 as follows

R1: (age = 50) ^ (Smoker = yes)) (Heart attack passtive= yes)

If the condition holds the true for a given tuple, then the antecedent is satisfied.

C. Comparative result



IV. Conclusion

In this paper, we compare t algorithms on dataset with some parameter. In Relevant dataset contain simple and class attribute. We have presented an intelligent and effective heart attack prediction methods using data mining. Firstly, we have provided an efficient approach for the extraction of significant patterns from the heart disease data warehouses for the efficient prediction of heart attack Based on the calculated significant weight age, the frequent patterns having value greater than a predefined threshold were chosen for the valuable prediction of heart attack. The goals are to be evaluated against the trained models. All these models could answer complex queries in predicting heart attack.

V. Future Enhancement

In our future work, this can further enhanced and expanded. For predicting heart attack significantly 15 attributes are listed. Besides the 15 listed in medical literature they can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of just categorical data. They can also use Text Mining to mine the vast amount of unstructured data available in healthcare databases.

References

[1] Han,J.,Kamber, M.:"Data Mining Concepts and Techniques", Morgan Kaufmann Publishers,2006.
[2] Margaret H.Dunham"Data Mining: Introductory and Advanced Topics".

[3] Jyoti soni, sunita soni e.,"predictive Data Mining for Medical Diagnosis: on overview of Heart Disease Prediction"; International Journal of computer Applications (0975-888) volume 17-No.10, June 2012.
[4] Ms.Ishtake S.H, Prof.Sanap S.A.," Intelligent Heart Disease Prediction System Using Data Mining Techniques"; International Journal of Health care & Biomedical Research, Volume: 1, Issue: 3, Pages 94-101. April 2013.
[5] Kiran Jyoti, Nidhi Bhatla Ludhiana, India"An Analysis of Heart Disease Prediction Using Different Data Mining Techniques"; International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol.1 Issue 8, October-2012**.**
[6] Abdulla.A.Aljumah, Mohammed Gulam ahamed, Mohammed Khubeb Siddiqui"Application of Data mining: Diabetes health Care in Young and Old Patients"; International Journal of computer and Information Science 25,127-136.October -2012.
[7] Wipada chanthaweethip, sumanata guha "Temporal Data Mining and Visualization for Treatment Outcome Prediction in HIV Patients"; Proceeding of the International Neural Network Society Winter Conference (INNS-WC). September- 2012.
[8] Shadab Adam Pattekari and Asma Praveen"Parediction System For Heart Disease Using Naïve Bayes"; International Journal of computer and Mathamatical Sciences ISSn 2230-9624.Vol 3,pp 290-294,2012.
[9] ShamsherBahadur,Patel,Y.S.Kumaraswamy,"Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4,August -2009.
[10] R.Bhuvaneswari and K.Kalaiselvi" Naive Bayesian Classification Approach in Healthcare Applications". ; International Journal of Computer Science and Telecommunications [Volume 3, Issue 1, January 2012].
[11] Muhammad Hariz Muhammad Adnan, Wahidah Husain, Nur'Aini Abdul Rashid," Data Mining for Medical Systems: A Review".; Proc. of the International Conference on Advances in Computer and Information Technology - ACIT 2012.
[12] Bekir Karlık and Emre Öztoprak, Member, IACSIT," Personalized Cancer Treatment by Using Naive Bayes Classifier"; International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012
[13] Peter W. F. Wilson, Ralph B. D'Agostino, Daniel Levy, Albert M. Belanger, Halite Silbershatz," Prediction of Coronary Heart Disease Using Risk Factor Categories", International Journal of American Heart Association, Circulation. 1998; 97:1837-1847.
[14] Karlık and Emre Oztoprak, Member, "Personalized Cancer Treatment by Using Naive Bayes Classifier"IACSIT Vol. 2, No. 3, [June 2012]
[15] Dr. Yashpal singh, alok singh chauhan," Neural Networks in Data Mining"; Journal of Theoretical and Applied Information Technology,2005-2009.
[16] K.Srinivas B.Kavitha Rani Dr. A.Govrdhan," Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks"; (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255
[17] K. Manjusha, K. Sankaranarayanan, SeenaP**,"**Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification**",** Volume 4, Issue 1, January 2014 ISSN: 2277128X International Journal of Advanced Research in Computer Science and Software Engineering.
[18] Chen, J., Greiner, R. Comparing Bayesian Network Classifiers. In Proc. Of UAI-99, pp.101–108, 1999.
[19] "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: New Mexico. 2001 – 2003.
[20] "Heart disease" from http://wikipedia.org.
[21] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Techniques, 978-1-4244-1968-5/08/$25.00 ©2008 IEEE.
[22] Richard N. Fogoros, M.D, The 9 Factors that Predict Heart Attack90% of heart attacks are determined by these modifiable risk factors, About.com Guide.
[23] Harleen Kaur and Siri Krishan Wasan, Empirical Study onApplication of Data Mining Techniques in Healthcare Journal of Computer Science 2 (2): 194-200, 2006 ISSN 1549-3636 © 2006 Science Publications.
[24] Bressan, M. and J. Vitria, On the selection and classification of independent features. Pattern Analysis and Machine Intelligence,IEEE Transactions on, 2003. 25(10): p. 1312-1317.
[25] Juan Bernabé Moreno, One Dependence Augmented Naive Bayes, University of Granada, Department of Computer Science and Artificial Intelligence.

