Heart Attack Prediction for Diabetic's Patient using Data Analytics

T Rajasekaran, R Tharani priya[*], V Pavithra, P Sudharshanam

Department Of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore-641607,Tamilnadu, India.

*Corresponding Author: R. Tharani priya

E-mail: tharanipriya900@gmail.com

Abstract

Good health is nothing but the condition where both our body as well as our mind is functioning properly. The main causes behind poor health conditions are diseases, improper diet, injury, mental stress, lack of hygiene, unhealthy lifestyle, etc. Over the past few years, our lifestyle has changed and often tends to ignore the importance of healthy living in one way or the other. Extreme cases in diabetic's leads to cardio vascular events, lipids, blood pressure, mortality. The calories for some eatables are collected which should be followed by the diabetics patient. This is implanted in weka tool .From the recommended low calorie food items diabetic patient can be reduced from the risk of heart attack. By following this recommended food items number of death rates can be reduced.

1. Introduction

In India, number of death is 2,596,993.Therfore death rate is 821.5 deaths per 100,000 populations. The life expectancy is 78.8 years. Infant Mortality rate is 5.96 deaths per 1,000 live births and number of deaths of heart disease is 611,105. India's no.1 killer disease is a heart attack. Heart attack is the death of a segment of heart muscle caused by the loss of blood supply **[2].** The blood supply is usually lost because a coronary artery, one that supplies blood to the heart muscle, has a blood clot, a blockage. If some of the heart muscle dies, the patient experiences chest pain and electrical instability of the heart muscle tissue. Heart attacks occur as a result of coronary heart disease (CHD). CHD is a condition in which a waxy substance called plaque develops inside of the coronary arteries. These arteries supply oxygen-rich blood to your heart.

This heart disease is due to cardio vascular events (heart attack), lipids, blood pressure. Etc and some other factors too. The paper is totally describing the diabetic's patient with uncontrolled food stuffs experiencing heart diseases. For effective diabetes management, blood glucose, blood pressure and lipid profile must be monitored

regularly **[6].** For maintaining all these factors they must aware of controlled foodstuff like less intake of carbohydrates, sugar etc.

1.1 Data Analytics

Data analytics (DA) is the science of examining raw data with the purpose of drawing conclusions about that information. Data analytics is used in industries to make better business decisions and in the sciences to verify or disprove existing models or theories **[7].**

Analysis refers to breaking a whole into its separate components for individual examination. Data analysis is a process for obtaining raw data and converting it into information useful for decision-making by users. Data is collected and analyzed to answer questions, test hypotheses or disprove theories **[13].**

2. Heart Attack

Heart attacks are the leading cause of death of men and women. It is important to know the warning signs because delaying treatment just a few minutes could result in death. Women are less likely than men to believe they are having a heart attack and consequently often delay seeking treatment **[12].**

The most common symptom for both men and women are discomfort or pain in the chest. However, women are more common than men to have other warning signs such as *pain or discomfort in the jaw or neck and shortness of breath, light-headedness, cold sweat or nausea.* Most risk factors are modifiable, such as sedentary life style, obesity, cigarette smoking, high cholesterol, high blood pressure, and diabetes. Other risk factors such as age and genetics (family history) cannot be modified **[11].**

3. Diabetics

Diabetes mellitus commonly referred to as diabetes, is a group of metabolic diseases in which there are high blood sugar levels over a prolonged period. Symptoms of high blood sugar include frequent urination, increased thirst, and increased hunger. If left untreated, diabetes can cause many complications. Serious long-term complications include cardiovascular disease; stroke, chronic kidney failure, foot ulcers, and damage to the eyes. There are three main types of diabetes mellitus: Type 1- Diabetes mellitus results from the pancreas's failure to produce enough insulin. Type 2- Diabetes mellitus begins with insulin resistance, a condition in which cells fail to respond to insulin properly. Gestational diabetes- Gestational diabetes is the third main form and occurs when pregnant women without a previous history of diabetes develop a high blood-sugar level.

3.1 Dataset for Calories

Here calories for some vegetables, fruits and sweets are collected. If diabetics patient had high calories it leads to heart attack .From the collected data low calorie food is allowed to a diabetics patient. The data is also collected diseases due to low and high calorie levels.

NAME	QUANTITY	CALORIES	EDIBILITY

APPLE	1MEDIUM	65	EDIBLE
PEARS	1MEDIUM	75	EDIBLE
APRICOTS	1MEDIUM	20	EDIBLE
BLUE BERRIES	1CUP	50	EDIBLE
KIWIFRUIT	1 MEDIUM	40	EDIBLE
GRAPE FRUIT	1 MEDIUM	20	EDIBLE
GRAPES	1 BUNCH	310	NEDIBLE
POTATO	1 MEDIUM	125	NON EDIBLE
CAULIFLOWER	1 CUP	30	EDIBLE
LETTICE	1 CUP	5	EDIBLE
CABBAGE	1 CUP	20	EDIBLE
SPINACH	1 CUP	15	EDIBLE
ONION	1 CUP	30	EDIBLE
CUCUMBER	1MEDIUM	10	EDIBLE
TOMATOES	1MEDIUM	20	EDIBLE
CARROT	1MEDIUM	55	EDIBLE
RADDISH	1CUPSLICE	19	EDIBLE
BARFI	1PC	100	NON EDIBLE
HALWA	1PC	570	NON EDIBLE
GULAB JAMUN	1PC	100	NON EDIBLE
JALEBI	1PC	200	NON EDIBLE
MYSORE PARK	1PC	357	NON EDIBLE
RASAGOLI	1PC	150	NON EDIBLE

Table 1: Data set for calories

3.2 Naive Bayes

The Naive Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given data set. The algorithm uses Bayes rule and assumes all attributes to be independent given the value of the class variable. This conditional independence assumption rarely holds true in real world applications, hence the characterization as Naive yet the algorithm tends to perform well and learn rapidly in various supervised classification problems **[9].**

Bayes's rule: $P(H \mid E) = P(E \mid H) \times P(H) \, P(E)$

The basic idea of Bayes's rule is the outcome of a hypothesis or an event (H) can be predicted based on some evidences (E) that can be observed. A priori probability of H or P(H): This is the probability of an event before the evidence is observed. A posterior probability of H or P(H | E): This is the probability of an event after the
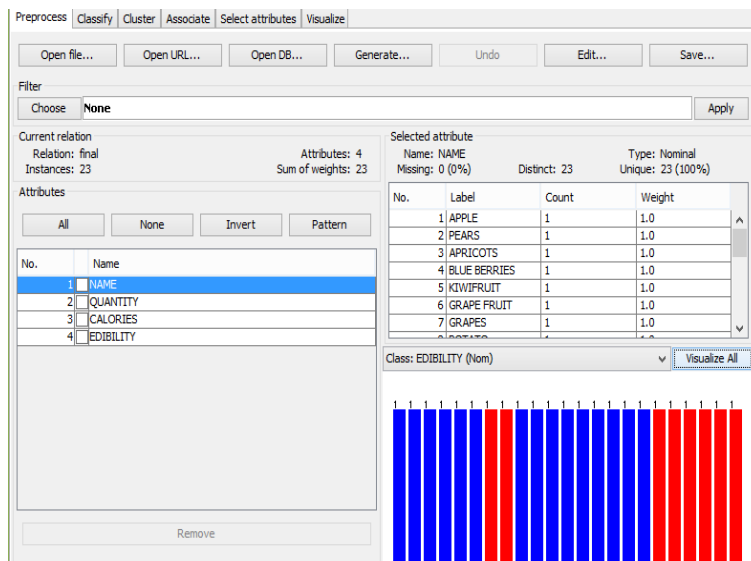
Figure 3.1.1: Preprocess for calories

For calorie data set the naive bayes gives 100% classification. It classifies the edible and non edible food items that can be eaten by diabetic's patient. It initially observes the calorie levels and later classifies edible and non edible stuffs by indicating in terms of p(H) and p(E) and by the above formula it gives 100% classification.
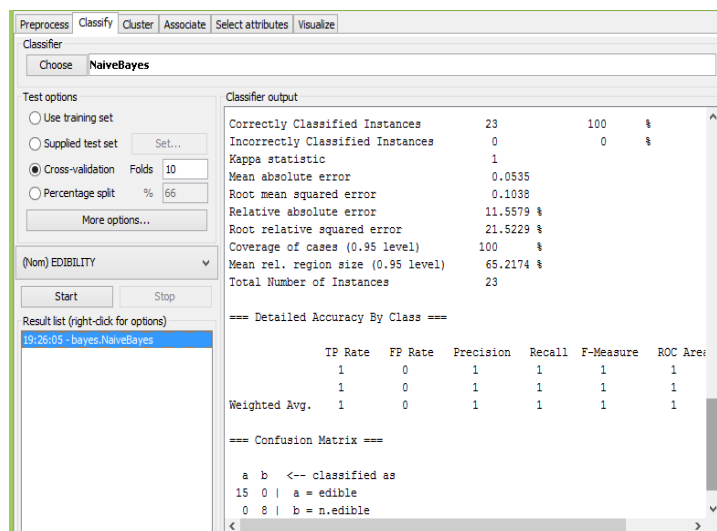

Figure 3.2.1: Classification using NaiveBayes

Figure 3.2.2: Tree representing edibility

Figure 3.2.3: Food items and its edibility

4. Diseases

Normal calorie limit for a human is 1200-2400. If calorie limit falls below 1200 it causes diseases like Gallstone, Osteoporosis etc. If calorie limit goes above 2400 it causes diseases like Diabetes, Heart Attack etc.

	A	B	C	D
1	risk level	diseases	prediction	
2	below1200	osteoporosis	unsafe	
3	below1200	gallstonstones	unsafe	
4	below1200	constipation	unsafe	
5	1200-1400	normal	safe	
6	above 2400	heart failure	unsafe	
7	above 2400	diabetics	unsafe	
8	above 2400	osteoarthritis	unsafe	
9				
10				
11				
12				
13				

Table 2: Data collection for diseases based on calories

4.1 J48 Algorithm

J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. C4.5 is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision trees generated by C4.5 can be used for classification. With this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple**[22][23][24].** Generally J48 is a good prediction algorithm. It predicts based on decision tree. The left most trees indicate diseases occurring due to low calories and right most tree indicates diseases occurring due to high calories and vice versa.
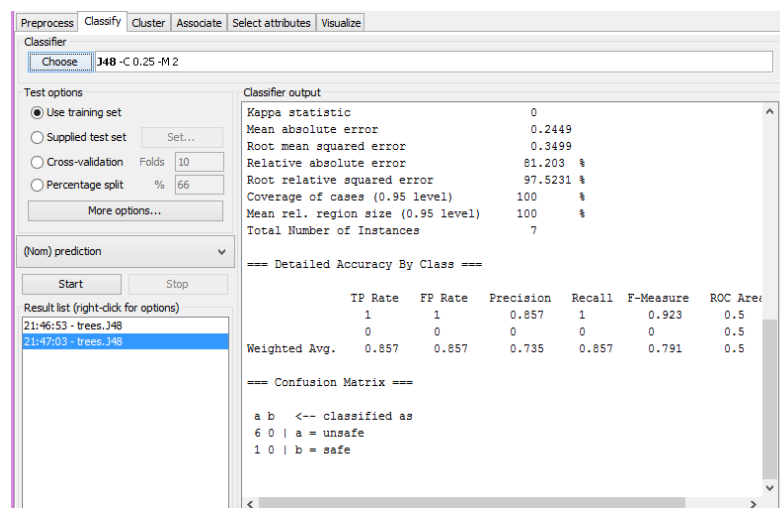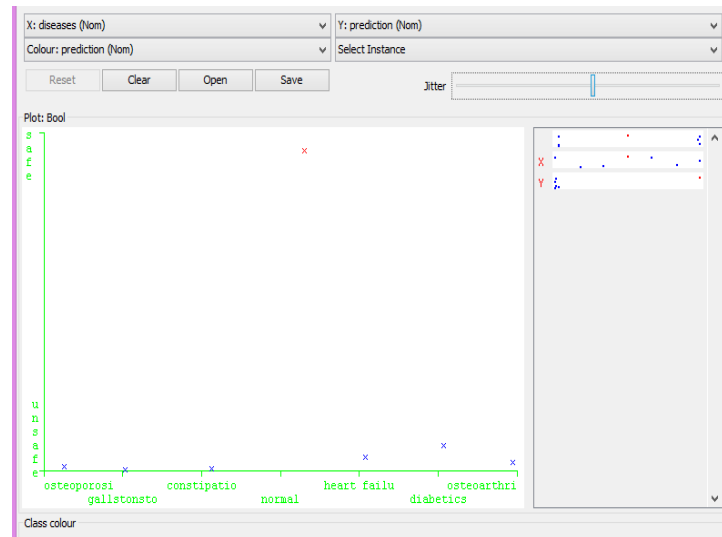

Figure 4.1.1: Classification using J48

Figure 4.1.2: Represents diseases and its prediction

5. Conclusion

In our work we have predict the chances of getting a heart disease using calories from diabetic's treatment and we have shown that it is possible to reduce heart disease vulnerability in diabetic patients with reasonable accuracy. Classifiers of this kind can help in early detection of the vulnerability of a diabetic patient to heart disease. There by the patients can be aware to change their lifestyle. This will result in preventing diabetic patients from being affected by heart disease, there by resulting in low mortality rates as well as reduced cost on health for the state.

6. Future work

Predicting the diseases like low bp, etc... That leads to heart attack. Another idea is to prevent the people from heart attack. Sensor is inserted in the mobile phone. When the person is in abnormal condition the sensor detects the condition of the person and senses the signal and send to the mobile phone, the light internally blinks in it causing alert messages to the nearby ambulance using shortest path algorithm. Using GPS (global positioning system) a particular person who is in abnormal condition can be tracked by the paramedic.

References

[1]   http://www.dailymail.co.uk/sciencetech/article-2408748/The-sensor-reads-heartbeat-feet-away-WITHOUT-touching-you.html
[2]   http://indiatoday.intoday.in/story/India%27s+no.1+killer:+Heart+disease/1/92422.html
[3]   https://en.wikipedia.org/wiki/Diabetes_mellitus
[4]   http://www.buzzle.com/articles/why-is-health-important.html
[5]   American Diabetes Association Diabetes Care http://care.diabetesjournals.org/content/33/12/e147.abstract
[6]   CII and Lilly release white paper on diabetes care at NCD Summit "Synergizing Efforts in Diabetes Care at the Tertiary Level"
[7]   Definition from whatIs.com searchdatamanagement.techtarget.com/definition/data-analytics

[8]   Judd, Charles and, McCleland, Gary (1989) Data Analysis. Harcourt Brace Jovanovich. ISBN 0-15-516765-0.
[9]   George Dimitoglou, James A. Adams, and Carol M. Jim," Comparison of the C4.5 and a Naïve Bayes Classifier for the Prediction of Lung Cancer Survivability"
[10]   O'Neil, Cathy and, Schutt, Rachel (2014)," Doing Data Science" O'Reilly, ISBN 978-1-449-35865-5.
[11]   G. Parthiban, A. Rajesh, S.K.Srivatsa, "Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method"
[12]   "saving lives, protecting people-heart attack ",Centers for disease control & prevention –CDC CDC 2417
[13]   Hort on works data platform, powered by apache hadoop-hadoop
[14]   Bennett, S. (1993).,"A History of Control Engineering 1930-1955. London: Peter Peregrinus Ltd.", on behalf of the Institution of Electrical Engineers. ISBN 0-86341-280-7-SENSOR
[15]   Jump up^ Bănică, Florinel-Gabriel (2012),"Chemical Sensors and Biosensors: Fundamentals and Applications",Chichester, UK: John Wiley & Sons. p. 576. ISBN 9781118354230.-SENSOR.
[16]   www.urbandictionary.com/define.php?term=AMBULANCE+DRIVER
[17]   www.ibm.com/software/data/bigdata/industry-healthcare.html
[18]   Global Healthcare and "Big Data" Redefining the Possible with Real-Time Information
[19]   "Prediction of Heart Diseases and Cancer in Diabetic Patients Using Data Mining Techniques", ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645,] Indian Journal of Science and Technology, Vol 8(14), IPL 018, July 2015
[20]   MS.TEJASHRI N. GIRI et al., "Data Mining Approach For Diagnosing Type 2 Diabetes" DATE OF PUBLICATION: JAN 02, 2015 ISSN: 2348-4098 VOL 2 ISSUE 8 NOV-DEC 2014
[21]   International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Web Site: www.ijettcs.org Email: editor@ijettcs.org Volume 4, Issue 1, January-February 2015 ISSN 2278-6856 Intelligent Heart Disease Prediction System With MongoDB
[22]   Jay Gholap Dept. of Computer Engineering College of Engineering, Pune, Maharashtra, India "PERFORMANCE TUNING OF J48 ALGORITHM FOR PREDICTION OF SOIL FERTILITY"
[23]   Margaret H. Danham,S. Sridhar, " Data mining, Introductory and Advanced Topics", Person education , 1st ed., 2006
[24]   Aman Kumar Sharma, Suruchi Sahni, "A Comparative Study of Classification Algorithms for Spam Email Data Analysis", IJCSE, Vol. 3, No. 5, 2011, pp. 1890-1895 [24]http://www.jstor.org/discover/10.2307/4039841