 ISBN : 978-93-5254-640-5

Prediction of Typhoid Disease using Naive Bayes Classifier

Dr. S. D. Mundhe[#1], Mr. D.R. Vidhate[*2]

Director[1], Sinhgad Institute of Management and Computer Application (SIMCA), Narhe,
Asst. Prof [2], College of Computer Application for Women, Satara,
[1]director_mca_simca@sinhgad.edu
[2]vidhatedhananjay@rediffmail.com

***Abstract*— Today huge amount of data is being collected for analysis of various diseases in the field of medical science. But unfortunately collected data are not mined. It is very important to mine the data for finding hidden information. The popular technique of information technology such as knowledge mining helps to discover hidden patterns. This paper focuses on study of typhoid disease. Typhoid infection is often passed on through bacteria or through contaminated food and drinking water, and it is more prevalent in places where hand washing is less frequent. It can also be passed on by carriers who do not know they carry the bacteria. It requires some tests to be carried out by the patients to detect the disease. This paper presents how naive bays classifier technique of knowledge mining can systematically study data of typhoid patients based on their symptoms and parameters suggested by doctors. This paper will be helpful for medical practitioners for effective decision making or decrease the number of tests for patients.**

***Keywords*— Knowledge mining, Naive bays classifier, Typhoid disease, decision making.**

I. Introduction

In India, typhoid is a common bacterial infection. Symptoms include a high fever and gastrointestinal problems. Some people carry the bacteria without developing symptoms. If a disease is not treated properly it is fatal in around 25 precent of cases. This paper helps to decrease laboratory test of patents. It will help medical practitioners to take intelligent decisions and also save their valuable time. The main aim of this paper is to study data of typhoid patients by using naïve bays classifier technique of knowledge mining which will help to diagnose the disease.

II. REVIEW OF LITERATURE

Margaret H. Dunham, S. Sridhar in his book "Data Mining Introductory and advanced topics" mentioned that, database is growing at phenomenal rate. The users are expecting more sophisticated information from database. Simple structured query language queries are not adequate to support the increased demands for information. Data mining is used to solve this problem by finding hidden information in a database.

T V Suresh Kumar, B Eswara Reddy, Jagdish S Kallimani in his book "Data Mining principles and applications" highlighted that, keeping in mind the changeable and complicated needs of business environment, it is necessary to examine the need of evolution in the traditional decision support techniques. The aim is to intensify the need for integrated performance measurement and management, which are currently based on historical data. Because of the nature of challenges and trends in the retail industry, it is considered to be an appropriate scenario.

Andrew Newmen & Peter Cullen in their book 'Retailing: Environment & Operations' have focused on various aspects of Retailing as a Business. They have considered retailing as a important part of our changing society and major source of employment. The retailing is closely tied to the changing moods of the consumers and new ways of business, spread on by the impressive development in Technology and Management Theory. The book provided importance to retailing, including Logistics and Distribution, Merchandising, Store Layout and design, pricing and location strategy. Retail services and out of store retailing have been included as new areas. This book tried to find out the different market structures that are required for retail operations. This helps the readers to understand different facets, challenges and changes that are happening in the retailing environment.

III. TYPHOID DISEASE

Typhoid fever is an acute infectious illness associated with fever that is most often caused by the Salmonella typhi bacteria. The bacteria are deposited through fecal contamination in water or food by a human carrier and are then spread to other people in the area. Typhoid fever is rare in industrial countries but continues to be a significant public health issue in developing countries. Typhoid is a bacterial infection that can lead to a high fever, diarrhea, and vomiting. It can be fatal. It occurs predominantly in association with poor sanitation and lack of clean drinking water. According to the most recent estimates (published in 2014), approximately 21 million cases and 222 000 typhoid-related deaths occur annually worldwide.

Typhoid fever symptoms are:-

- poor appetite
- headaches
- diarrhoea
- generalized aches and pains
- fever
- lethargy
- weakness

Countries with less access to clean water and washing facilities typically have a higher number of typhoid cases. This can happen through an infected water source or when handling food. Even when the symptoms of typhoid have passed, it is still possible to be carrying the bacteria. The only

ISBN : 978-93-5254-640-5

effective treatment for typhoid is antibiotics. The most commonly used medicines are ciprofloxacin and ceftriaxone. Some studies have found Salmonella typhimurium resistance rates to be around 35 precent. No animals carry this disease, so transmission is always human to human. For those traveling to high-risk areas, typhoid vaccines are now available.

IV. KNOWLEDGE MINING

Knowledge mining is process of analysing data from different perspectives and summarizing it into useful information. Different knowledge mining services can help us increase revenue. Knowledge mining is used to construct six types of models aimed at solving business problems: classification, regression, time series, clustering, association analysis, and sequence discovery. The first two, classification and regression are used to make predictions, while association and sequence discovery are used to describe behaviour. Clustering can be used for either forecasting or description. Time series analysis and forecasting are used to detect temporal patterns from historical time-dependent data and project the detected patterns into the future. Time series analysis, like all other forms of data analysis, is used to characterize or explain the reasons for the behaviour of a system and to predict its future behaviour.

V. NAÏVE BAYES CLASSIFEIR

Bayes' theorem is stated mathematically as the following equation.

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

Where A & B are events & P(B)#0

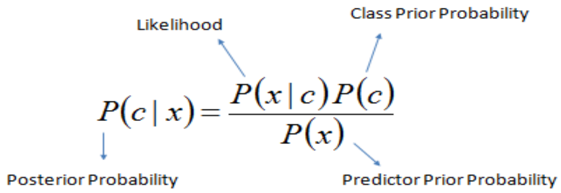
P(A/B) is a conditional probability the likelihood of event A occurring given that B is true.

P(B/A) is a conditional probability the likelihood of event B occurring given that A is true.

P(A) and P(B) are the probabilities of observing A and B independently of each other. This is known as marginal probability.

The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

P(c|x) is the posterior probability of class (target) given predictor (attribute).

P(c) is the prior probability of class.

P(x|c) is the likelihood which is the probability of predictor given class.

P(x) is the prior probability of predictor.

Example:- Following is data collected from the hospital for study of various patients suffering from typhoid disease. Their symptoms and diagnosis are shown in the following table.

Headache	Fever	Poor Appetite	Generalized pain	Weakness	Typhoid
Y	N	Y	Y	Y	Y
Y	N	N	Y	Y	N
N	N	Y	N	Y	N
Y	Y	Y	N	N	Y
N	Y	N	Y	N	N
N	Y	Y	N	N	Y
N	Y	N	Y	Y	Y
Y	Y	N	Y	Y	Y

Table No: 1

If there is new entry of patient in the hospital having following symptoms which listed below in the Table no2. So do we believe that a patient with following symptoms has the typhoid or not?

Table No: 2

Headache	Fever	Poor Appetite	Generalized pain	Weakness	Typhoid
Y	N	Y	N	Y	?

The stepwise algorithm for naive bays classifier is mentioned below:

Step1: Start

Step2: Accept input attribute

Step3: for each input attribute find its probability against all output

Step4: Multiply all the probabilities with respect to each attribute

Step5: Select probability with maximum value

Step6: Attribute will belong to class variable with maximum value

Step7: End

By using above algorithm we can conclude that patient has been suffered from typhoid or not. This will be tested from the incoming attribute which belongs to class variable with maximum value.

First all possible individual probabilities are computed based on target attribute of typhoid contained all probabilities of attribute of typhoid disease.

P(typhoid =Y)= 5/8= 0.625
P(heachache=Y/typhoid =Y) = 3/5=0.6
P(typhoid =N)=3/8=0.375
P(heachache=Y/typhoid =N) = 1/3=0.3333

Just like above we can simply compute the possible probabilities for all conditions and this probabilities are enlist in table no: 3

Table No: 3

P(Typhoid=Y)	0.62	P(Typhoid=N)	0.37
P(heachache=Y/typhoid=Y)	0.6	P(heachache=Y/typhoid=N)	0.33
P(heachache=N/typhoid=Y)	0.4	P(heachache=N/typhoid=N)	0.66
P(fever=Y/typhoid=Y)	0.8	P(fever=Y/typhoid=N)	0.33
P(fever=N/typhoid=Y)	0.2	P(fever=N/typhoid=N)	0.66
P(poor appetite=Y/typhoid=Y)	0.6	P(poor appetite=Y/typhoid=N)	0.33
P(poor appetite=N/typhoid=Y)	0.4	P(poor appetite=N/typhoid=N)	0.66
P(Generalizd pain=Y/typhoid=Y)	0.6	P(Generalizd pain=Y/typhoid=N)	0.66
P(Generalizd pain=N/typhoid=Y)	0.4	P(Generalizd pain=N/typhoid=N)	0.33
P(weakness=Y/typhoid=Y)	0.6	P(weakness=Y/typhoid=N)	0.66
P(weakness=N/typhoid=Y)	0.4	P(weakness=N/typhoid=N)	0.33

And then we decide that p has split into two cases one for Y and second for N
p1->
argmax P(typhoid=Y)*P(heachache=Y/typhoid=Y)* P(fever=N/typhoid=Y)* P(poor appetite =Y/typhoid=Y)* P(Generalizedpain=N/typhoid=Y)*P(weakness=Y/typhoid =Y)

P1=0.62*0.6*0.2*0.6*0.4*0.6
P1=0.0107

P2->
argmax
P(typhoid=N)*P(heachache=Y/typhoid=N)*P(fever=N/typhoid=N)*P(poor appetite =Y/typhoid =N) *P(Generalized pain=N/typhoid =N)* P(weakness=Y/typhoid =N)

P2=0.375*0.33*0.66*0.33*0.33*0.66
P2=0.0058

Therefore, the argument probability of p2 seems smaller than p1 so that newly entered patient in the hospital is suffering from typhoid diseases.

VI. CONCLUSION

Naive bays classifier technique of knowledge mining can be used for the diagnosis of typhoid disease. This paper has shown a simplest way to identify the typhoid patients based on the data of typhoid patients. This result has shown that knowledge mining could be effectively utilized for diagnosis of various diseases in the field of medical science.

VII. FUTURE RESEARCH

Knowledge mining techniques such as classification, clustering, association rules and many other techniques could be effectively utilized for studying patient data. In future, it is possible to get more accurate diagnosis of typhoid patients by increasing various parameters suggested by doctors by using knowledge mining.

REFERENCES

[1] Margaret H. Dunham, S. Sridhar, "Data Mining: Introductory and Advanced Topics" Pearson Education, Inc., 2006.

[2] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20thVLDB conference, pp 487–499.

[3] http://www.itl.nist.gov/div898/handbook/prc/section1/prc13.htm

[4] Briony J Oates , Researching Information Systems and Computing, South Asian Edition, 2006.

[5] Dimitris Papakkiriakopoulos, "Predict on-shelf product availability in grocery retailing with classification methods," Expert Systems with Applications, Volume 39, Issue 4, PP. 4473-4482, March 2012.

[6] Nisha Rathore, "A Study on Consumer Behavior Towards Organized and Unorganized Retailing", Journal of Research in Commerce and Management, Volume No.1, Issue No.8, 2012.

[7] James Cho, Gregory S Ching, Thai-Ha Luong," Impulse buying behavior of Vietnamese consumers in supermarket setting", International Journal of Research Studies in Management, Vol 3, No 2, 2014.

[8] Niren Sirohi , Edward W. McLaughlin , Dick R. Wittink, "A model of consumer perceptions and store loyalty intentions for a supermarket retailer", Journal of retailing, 74, 2 (1998), 223-245.

[9] Karen A. Machleit, Susan Powell Mantel, "Emotional response and shopping satisfaction Moderating effects of shopper attributions", Journal of Business Research 54 (2001) 97– 106.