ISSN: 0975-766X
CODEN: IJPTFI
Available Online through ***Research Article***
www.ijptonline.com

NOVEL APPROACH FOR CLASSIFYING DISEASE-PREVENTION MECHANISM USING NAIVE BAYES ALGORITHM [ELECTRONIC-HEALTH CARE SYSTEM]

[*]Priya Govindarajan, [1]Michael Raj.T.F, [2]Priyanka.M, [3]Bharathi.C, [4]Vishnu Priya.K
[*]Asst. Prof., SASTRA University, Kumbakonam.
[1]Asst. Prof., SASTRA University, Kumbakonam.
[2,3,4] PG Student, SASTRA University, Kumbakonam.
Email: priya@src.sastra.edu

Received on: 19.10.2016 *Accepted on: 20.11.2016*

Abstract:

Objective: The main objective of this work is to improve the time complexity, speed and quality of classification, when compared with the existing classification methodologies.

Method/Statistical analysis: Electronic [E]-Healthcare system is a web application which creates awareness about fatal and non-fatal diseases by the parameters provided by the end-user. The system predicts a disease and its corresponding prevention mechanism based on a group of symptoms given by the end-user. The proposed system takes into account the vital factors such as age, blood pressure and sugar level for indicating a particular disease and its corresponding prevention mechanism. This novel approach of classification is manipulated through Naïve Bayes algorithm.

Findings: The disease is predicted through its occurrence and count. After which the manipulation is carried off to drag the final probability, the disease with highest (max. value) probability is the resultant (predicted) disease.

Application/Improvements: Most of the drawbacks of existing classification methodologies were improved through feature selection. The developed web application can act as a subordinate system rather than a full-fledged system.

Keywords: Health Care System, Naive Bayes Algorithm, Disease, Classification method, Prevention.

Introduction: Most of the available data, content, information is mostly propagated into digital form. In order to understand the text information in depth some process [1, 2, 3] are to be performed. The classification [2,4] involves finding rule that portions the data into disjoint groups. The aim of classification is to construct a group of modules that can exactly identify the class of the various objects [5,6]. Text classification can be done by Decision Tree, Support vector

machine and wavelet neural network. Decision tree learning is a method for approximating discrete-valued target functions. Some leaf node provides the classification about the instance. A descendant of the root node is then created for possible value. Each internal node is labeled with the attribute it tests and its branches are labeled with name.

Classification [1] is done by Support Vector Machine (SVM). SVM is closely related to other fundamental classification algorithm. Support vector machine [4] training algorithm builds a model that will predict whether a new example falls into new category or the other. This is improved after decreasing the dimensions of data without the data loss. A linear model is a model that uses the linear combination of feature values.SVM searches to separate the positive and negative examples with maximum difference. Web based document classification is implemented by Neural network [4]. These are a new class of networks [7] which have been used with success in a wide range of applications. Wavelet neural network can classify the web information by non linear function approach that is pattern classification.

A drawback for this existing classification is that it's Complex or time consuming [1, 2, 8], an action and data reduction procedure can result in the loss of classification information. The feature selection improves both the speed and the quality of classification, by which one can overcome the existing drawback.

2. Proposed Methodology:

2.1. Naïve Bayes Algorithm:

Classifications are carried off with a simple technique called Naïve Bayes [2, 4, 9]. Probability can be manipulated by implementing Naïve Bayes classifiers in the methodologies of machine learning. Naïve Bayes can be used to calculate values in most of the practical applications [6,10]. It utilizes maximum likelihood methodology, without calculating the Bayesian probability. The value of each and every feature [3] differs from the other and this is considered to be the common principle of Naïve Bayes classifier. For example, Stroke is considered to be a disease if the user has the symptoms of muscle weakness, chest pain, etc... Even though each and every symptom has its own feature in classification, it is being grouped into a single disease called stroke.

2.2. Working Principle:

Resultant Disease=Possible occurrence/Count

Here, possible occurrence can be calculated by(probability + combination) in which combination refers to disease with symptoms ,here Naive Bayes algorithm plays a important role where it match a symptom with symptoms in the dataset.

Whereas probability can be calculated by dividing number of resultant disease with total number of disease. In which, combination value gets added with probability. Count can be calculated by adding the count of particular disease with the number of symptoms for one particular disease in the dataset [Table1].

Table1: Depicts a sample data set.

NO.	SYMPTOMS1	SYMPTOMS2	SYMPTOMS3	SYMPTOMS4	SYMPTOMS5	DISEASE
1.	Fever	Headache	Sensitivity to light	Vomiting	Red eyes	Yellow Fever
2.	Cough	Fever	Shortness of breath	Headache	Sweating	Pneumonia
3.	Tongue swelling	General fatigue	Shortness of breath	Dizziness	Weakness	Iron Deficiency anemia
4.	Tiredness	Thirst of dehydration	Irritability	Weight loss	Reduction of blood pressure	Juvenile diabetes
5.	Increased	Burning eyes	White discharge from eyes	Redness in inner eyelid	Itchy eye	Pink eye
6.	Weight loss	Heartburn	Pain in stomach	Vomiting	Acid Reflux	Ulcer

2.3. Training Example:

Sample Symptoms to Check:

1. Fever
2. Dizziness
3. Sweating

Here, the provided symptoms will be matched with all the symptoms given in the dataset. If both the data are matched, we can calculate the probabilities and multiply them. From the given sample 1 (i.e.) Fever is matched with all the symptoms given in the datasets and the combination is done. Similarly symptom2 and symptoms3 are matched with all the symptoms. For each disease- 3 probability values will appear, these three values are multiplied with count (i.e.) 1/number of disease. Finally calculation is done for all the disease given in the dataset. Naive Bayes algorithm takes the maximum value of a disease based on which the corresponding disease is being projected.

2.4. Calculation of the Fetched Symptoms:

1: Yellow Fever

(1)Fever/Yellowfever (2) Dizziness/ Yellowfever (3) Sweating/ Yellowfever

N=3 N=3 N=3

Nc=1 Nc=0 Nc=0

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=>(0)

2: Pneumonia

(1)Fever/Pneumonia (2) Dizziness/ Pneumonia (3) Sweating/ Pneumonia

N=3 N=3 N=3

Nc=1 Nc=0 Nc=1

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=> (0.193)(0.026)(0.193)

3: Iron Deficiency Anemia

(1)Fever/Iron Deficiency (2) Dizziness/ Iron Deficiency (3) Sweating/ Iron Deficiency

Anemia Anemia Anemia

N=3 N=3 N=3

Nc=0 Nc=1 Nc=0

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=>(0.026)(0.193)(0.026)

4: Juvenile Diabetes:

(1)Fever/ Juvenile diabetes (2) Dizziness/ Juvenile diabetes (3) Sweating/Juvenile diabetes

N=1 N=1 N=1

Nc=0 Nc=0 Nc=0

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=> (0.026)(0.026) (0.026)

5: Pinkeye:

(1)Fever/ pinkeye (2) Dizziness/ pinkeye (3) Sweating/pinkeye

N=1 N=1 N=1

Nc=0 Nc=0 Nc=0

.193)(0.026)(0.026)

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=> (0.026)(0.026)(0.026)

6: Ulcer

(1)Fever/ ulcer (2) Dizziness/ ulcer (3) Sweating/ulcer

N=1 N=1 N=1

Nc=0 Nc=0 Nc=0

M=5 M=5 M=5

P=1/6 P=1/6 P=1/6

ANS=> (0.026)(0.026)(0.026)

The maximum value occurs in pneumonia, so the resultant disease is pneumonia. It is described in [Table 2].

Table-2. Final Probability.

S.NO	NAME OF THE DISEASE	MANIPULATION	FINAL PROBABILITY
1	Yellow Fever	0.026*0.193*0.026*0.16	0.0000208749
2	Pneumonia	0.026*0.193*0.193*0.16	0.0001549558
3	Iron Deficiency anemia	0.026*0.193*0.026*0.16	0.000028749
4	Juvenile diabetes	0.026*0.026*0.026*0.16	0.000028122
5	Pink eye	0.026*0.026*0.026*0.16	0.000028122
6	Ulcer	0.026*0.026*0.026*0.16	0.000028122

3. Outliner Analysis:

It is hard to make a decision that a classification algorithm is efficient than other because it may works correctly with certain data sets. Resultant on performance of a classification algorithm is usually depends on its accuracy. We have used a set of training examples and then have tested it. We have 100 disease and 500 symptoms to predict the accurate disease name that the person has been affected. From which a sample data set is projected in [table 1].

Here, for example we include 6 disease and 30 symptoms to predict the proper identification of disease [table1]. The three Symptoms should be mandatorily given by the user, which is matched with the above 30 symptoms one by one. (i.e.) the first disease given in the dataset is yellow fever, the symptoms is matched with the symptoms in the disease yellow fever, if the symptoms are present in that particular disease number of combination will be increased to 1. If the combination does not occur it will be turned to 0.

For example, the given symptom fever, is first matched with yellow fever and if the symptom fever is present in yellow fever, then it will be calculated as Nc+p/n+m p is probability, Nc is combination (i.e) fever/yellow fever, M is total number of symptoms in one disease, n is count of disease. For above Example, yellow fever is unique among all the 6 disease so n value is one and there are five symptoms for each disease so m=5, number of combination is (fever/yellow fever).These combination is present only once so Nc=1.So it can be calculated as 1+0.16/5+1=0.193. Same process is repeated for all the three symptoms until it is being compared with all the 6 diseases, which is present in the sample data set [table1].An application [E-Health care system] using those above manipulations are being projected in [figure 1,figure 2]

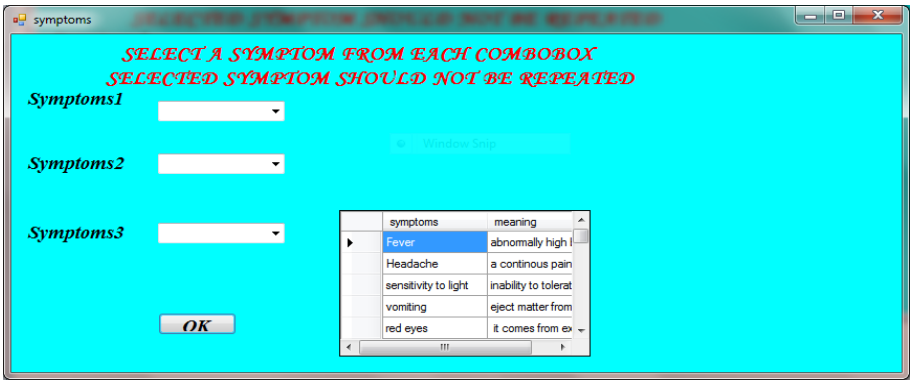

Figure1: Depicts a page with a collection of symptoms.

Figure-2: portrays the disease and its prevention mechanism for the specified symptoms.

4. Conclusion:

E-Healthcare system creates awareness about the fatal and Non-fatal diseases.This system helps to identify the diseases, based on the provided symptoms. According to the given symptoms, the disease may vary in its perspective. The goal of Healthcare is to predict the disease that the user has been affected. The causes and prevention measures are given to the end-user, based on the prediction of the disease. Not like any other existing system in which the user moves to the hospital to find the disease that they have been affected with, where as in the proposed health care system, we are identifying the disease rather than moving to a hospital. It's most useful for the user to check their health condition from anywhere.

References:

1. Kim S B, Han K S, Rim H C, Myaeng S H.Some Effective Techniques for Naive Bayes Text Classification ,IEEE Transactions on Knowledge and Data Engineering, 2006,18(11), pp.1457-66.
2. Ren J, Lee S D, Chen X, Kao B, Cheng R, Cheung D. Naive Bayes Classification of Uncertain Data, Ninth IEEE International Conference on Data Mining,Washinton, USA, 2009, pp.944-49.
3. Moura G, Roisenberg M. Probabilistic Fuzzy Naive Bayes, Brazilian Conference on Intelligent Systems, Natal, 2015, pp.246-51.
4. Leng C, Wang S, Wang H. Learning Naive Bayes Classifiers with Incomplete Data, International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, 2004,4, pp.350-53.
5. Ankalaki S, Majumdar J .Leaf identification based on back propagation neural network and support vector machine, Cognitive Computing and Information Processing international Conference on Noida, (CCIP), 2015, pp.1-7.
6. Govindarajan P, Ravichandran K S A. Framework For Identifying Disease-Treatment Relationship Using An Updated Database, International Journal of Applied Engineering Research, 2015, 10(16), pp. 36802-805.
7. Govindarajan P, Ravichandran K S. Data mining- an evolutionary arena, Research Journal of Applied Sciences, Engineering and Technology, 2014, 7(22),pp. 4749-53.
8. Govindarajan P, Ravichandran K S A .Combined Schema for Surveillance, Research Journal of Information Technology, 2014, 6(4), pp.406-12.

9. Michael Raj T F, Prasanna S. Implementation of ML Using Naïve Bayes Algorithm for Identifying Disease-Treatment Relation in Bio-Science Text, Research Journal of Applied Sciences, Engineering and Technology, 2013, 5(2),pp.421-26.
10. Vedanayaki M. A Study of Data Mining and Social Network Analysis, Indian Journal of Science and Technology,2014 Nov, 7(S7), pp.185-87.