ISSN:0975-9646

Data Analysis and Prediction of Hepatitis Using Support Vector Machine (SVM)

C. Barath Kumar[1], M. Varun Kumar[2,] T. Gayathri[3], S. Rajesh Kumar[4]

[1]MCA, [2]Assistant Professor, [3]M.S. Software Engineering, [4]MCA
VIT University, Vellore, Tamilnadu, India.

Abstract: - The project titled "Data Analysis and Prediction of Hepatitis Using Support Vector Machine (SVM)" is used for monitoring and predicting the Hepatitis level of the patients. Here we are using a machine based technology called Support Vector Machine (SVM). But the drawback is, we can't assure all the available and new data's are correct and related with each other. So, to find the hidden relationship's, removal of trivial data and noise avoidance feature before designating it, For all the mentioned process, we are using a method name called Wrapper method. It's mainly used to remove all the non-essential records and to establish the finite and accurate result. Our ultimate aim is to increase the accuracy level. In this paper we are using Rapid Miner as our tool. It is a Data Mining tool, which helps us to collect all the prior information from the patient's with their current Hepatitis level.

Keywords: - Hepatitis, Support Vector Machine (SVM), Wrapper Method, WEKA, Rapid Miner.

I. INTRODUCTION

Now a days the technology have been changed to machine based learning. Machine learning technology is one among the techniques of Artificial Intelligence (AI). This technique mainly helps to assign the patterns and spontaneous exchange between the aspects or attributes which we produced. The drawback originate during this techniques are the record set having more number of samples with respect to same aspect values and distributions. Because of this same aspect values, it causes many unwanted problems, disorganized and ineffective record set and leads to noisiness. Using the data analysis and prediction methodology will reduce the same aspect results and maintains the data efficiency. So that we can increase the accuracy level. Our ultimate aim is to anticipate the life time of the Hepatitis patient based on their medical data. The algorithm which we are using here is Support Vector Machine and it will predict the same aspect result and noisiness to maintain the accuracy in Hepatitis level of the patient. After eliminating or avoiding the same aspect records using SVM method, we are analyzing and measuring the result alone for finding the accuracy. The solutions to maintain the accuracy is mentioned in the proposed system elaborately.

II. EXISTING SYSTEM

The main problem featuring is to find and judge the originality among the record set. Before the modern technology has released the medical specialists correlating the patient's data with their old data or factual records or manually scribbled data to evaluate the disease level [1]. Under gone this type of activities may lead to awkward results and erroneous decisions. The record set which we collected already form different patient's will have different weightage in all the attribute but it has some same weightage also. For example, two same aged persons having Hepatitis and also their medical results inter- related in some attributes [5]. The manual evaluation will check the main set of related data's and if it is not related according to the difference in weightage they will calculate the level of the disease [7]. The Record which we have already contains many noisy attributes and it will leads to lower accuracy prediction of algorithms, if we applied or implemented it directly.

III. PROPOSED SYSTEM

The proposed system handles different techniques with different attributes and embed it with our proposed algorithms to maintain accuracy. The proposed technique which we adapted are mentioned below. According to the techniques only the process is involved here.

Techniques

The problem in detecting the fragment is distributing the attributes which involves in the original record set. So that the proposed featuring algorithm will endeavor on the data's which contains all the possibilities and that leads in superlative possible accuracy level.

A. Support Vector Machine (SVM)

Support vector machine is a popular biological machine and also one among the machine learning techniques. It is a computerized algorithm. The main process of this technique is, it organizes the label to its objects. It will recognize the same data in large collection of record set and also we can use this technique for pattern recognition. It solve and handles all the problems in scientific way or algorithmic way. SVM method excludes the inadequate elements by selecting the maximum surplus data that can separate the specific values from ambiguous values. But this technique needs each and every data in the record set must be mentioned in the matrix and numerical format. It won't allow the non-numerical data's. The matrix format will cross check the data set.

B. Chi- Squared Test

It is also a machine learning technique and it is also referred as (n^2) test. This test will help us to find the values of the attribute by considering the class it involved. There are many steps involved in this process. The steps are, it will maintain the weightage level, it will take hidden or missing values as a specific values. So that it will not emerge with original data. Finally it will take the numerical values and change them to binary to predict the result quickly. It is of statistical data and compare each data with its similar attribute. It detects the same values using Chi-Square Automatic Interaction Detector, simply called CHAID

C. Wrapper Method

Wrapper method is a method that involves the interpretation search with in the subset. This search contains pre-defined featured subsets as one side and evaluates the other or new incoming subsets. It wraps around the algorithm and pretend the relevant data. The process behind this wrapper method is the inadequate results that will find the accuracy in separate Bias. In this all the possibilities are treated as Ingredient selection problem. This method mainly used to filtering the results and forming the proper results for maintaining accuracy.

IV. Procedure

Based on the above techniques and record sets, we will see the procedural aspects for the following functions.

A. Analyzing the data

Rapid Miner is a tool which we are using here for analyzing the data. This software embeds with matrix, graphs, tables, charts and etc. Here the software uses the matrix which will compare the values of the record with the other record. The record mainly contains the attributes and its features. So, while comparing one record set with other record set, it establish the relationship among the data set and the same values predicted in to same record set and the different values are classified in to different record set. But the same values are not merged with each other and also it won't get involved in other patient record set. Here we are using a method called WEKA system which is done by wrapper method. It is nothing but a collection of all machine learning algorithms for distributing and predicting which is already done using Support Vector Machine (SVM)

B. Data Preprocessing

Using the evaluation of Chi- Squared test, the effective set of records improves the level of finding the accuracy. The records which involves in the mining techniques such as wrapper method and Support vector machine are to maintain the efficiency in the algorithms. The minimum feature will attain the highest accuracy and it is chosen as effective record set. The ultimate goal is to reduce or eliminate all the unwanted or irrelevant clinical records. This will leads in to appropriate and accurate result when compared with other techniques. Finally after the adequate selection of the all given records, the following are the reduced attributes, where the level of accuracy is increased. And based on the Table 1, we can come to the solution, how the attributes are generalized with their weightage.

C. Attribute Identity

The patient have to undergone various different attributes. Then only we come to know whether the patient is having the Hepatitis or not. If it is there, we have to identify and evaluate the attributes which are given in the Table 1, Attributes are calculated and applying the algorithm and generating the life time of the Hepatitis patient. Attributes can be identified in different aspects from the patient's body. Applying the algorithm will maintain the inadequate selection of data.

Attributes	Weightage
Age	Numbers
Albumin	1.232
Ascites	0.912
Bilirubin	0.991
Class	1.322
Histology	0.445
Malaise	0.552
Protime	1.0
Sex	Male (m) or Female (f)
Spiders	0.702

Table 1 – Attributes after the Adequate Selection

V. Experimental Results

The below accuracy chart (Table 2) will explain about the accuracy level of present and past in order to the given record set. Using the WEKA (Collection of all algorithms) system, the results are adequate and as well as accurate. The first results are calculated using the Hepatitis data using the SVM Library files only. After the chi square test and wrapper selection, the results calculated with only 7 attributes among all the 25 attributes which we had given as an input. And finally the accuracy level is increased up to 83%.

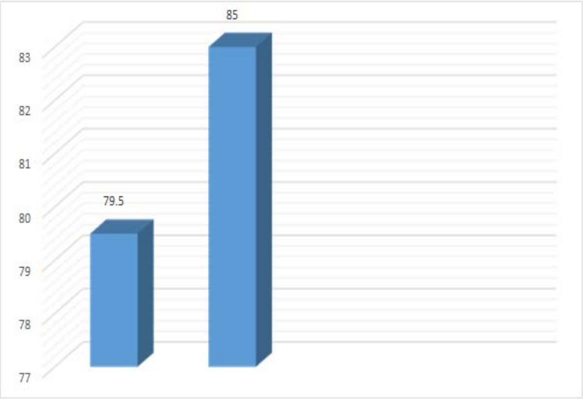

Fig 1- Accuracy Chart

The accuracy is increased here because the wrapper method and SVM method reduces the noisiness and finds the same value attributes or record set and categorized such aspects in separate table and compared those attributes in WEKA for obtaining the finite result. It have been evident that the adequate result increases the level of accuracy in all aspects with considering different attributes. WEKA test increases the accuracy.

VI. CONCLUSION

By considering all the above, We can come to a conclusion that, the Support vector machine (SVM) and WEKA method will join together and maintain the accuracy of the Hepatitis level. It is increased while comparing with the previously designed techniques and the adequate data reduces the noisiness and irrelevant data. The data mining techniques are used for maintain the bio medical data. Finally the result attained according to the increasing in accuracy level. And hence it is proved in Accuracy Chart.

REFERENCES

[1] Correlation based feature selection for machine learning (2009). A.H. Mark, Department of computer science. University of Waikato

[2] Estimating the confidence interval for prediction errors of support vector machine classifiers (2008). J. Bo and Xuegong, Journal of machine learning research.

[3] Prediction of Hepatitis prognosis using Support vector machine (2010). A.H. Rosline and A. Noraziah, Kuantan, Pahang.

[4] Feature Extraction for the Prediction of Liver Fibrosis Stages in Chronic Hepatitis (2007). C.Akifumi Miyazakiy, Miho Ohsaki, Eri Taniguchi.

[5] Hepatitis data Analysis and prediction model discovery using rapid miner (2008). Jinchao Han, Mohsen Behehsti, California Statement University.

[6] Data Mining: Concepts, Models, Methods and Algorithms (2002). Kantardize .M New jersy.

[7] www.cs.waikato.ac.nz for WEKA references

[8] The research and evaluation of Hepatitis metabolic function based on support vector machine (SVM) (2010). Chunquan, Shanghai, China

[9] Application of SVM in electromyography pattern recognition (2007). Chinese Journal of Sensors and Actuators.

[10] Experimentally optimal v in support vector regression for different noise models and parameter settings (2004). Athanassia C, Bernhard S, lkopf, Neural Network.