Open access Journal International Journal of Emerging Trends in Science and Technology

Analysis of J48 Algorithm in Classification-Ebola Virus

Authors

J Uma Mahesh[1], K.V.Naganjaneyulu[2], P. Likitha[3], K.N.S. Sindhu Aishwarya[4]

[1]Faculty of CSE, Assistant Professor, BIET

[2]Faculty of CSE, Professor, BIET

[3,4]Student, Dept of CSE, BIET

Emails: uma.mahesh@biet.ac.in, kvnaganjaneyulu75@gmail.com, likithapasumarthy@gmail.com, sindhukotha24@gmail.com

Abstract:

Data mining is the rising research area to overcome various queries and classification. It is one of the major problems in the area of data mining. In this article, we used J48 algorithm of the WEKA interface. It can be utilized for testing various datasets. J48 algorithm is based on C4.5 decision based learning. Weka is a collection of machine learning algorithms for data mining tasks. The algorithm can be applied directly to a data set or called from your own java code. Ebola virus is affected to the human kind from the affected animals. Proper care must be taken while contacting with the blood or bodily fluids of the affected animals. Proper care must be taken in case of clothing and cooking meat. Fruit bats are the main carriers of this virus. This disease has high risk of death. In this paper publication, the symptoms indicating Ebola have been analyzed in order to identify and cure the victims of the Ebola virus.

Keywords: *Data Mining, WEKA, J48, Ebola virus, Decision Tree,*

1. Introduction

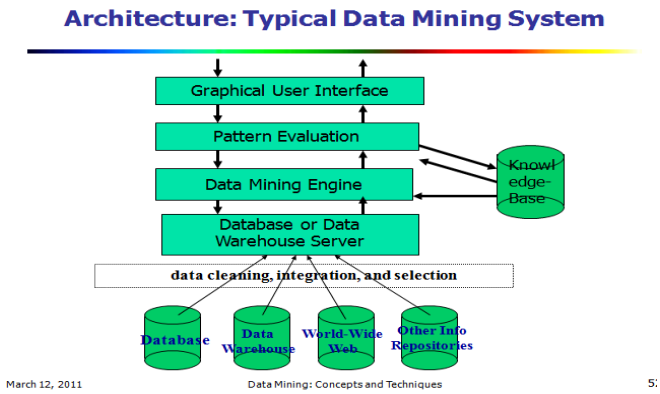
Data mining is the process of analyzing the patterns from a large number of datasets by joining methods from statistics and artificial intelligence with the application of database management. It is a rising area nowadays in good discipline. The idea of data mining is not to impose any rules on the datasets, but it helps us to guess and predict with a certainty analyzing only a very small set of data. Recently, data mining has gained a great attention in the knowledge and information industry due to the big availability of huge amounts of data and the upcoming need for converting that data into knowledge and meaningful information. The data mining technology is very reliable for database technology, statistical analysis, artificial intelligence and it has a great commercial value. It has a wide penetration in the areas such as retail, insurance, telecommunication, and power industries.

There is a need of good system architecture for the data mining to make use of the software environment. Data mining tasks are very efficient and effective way to exchange information with other systems. Data mining helps us to predict the future with the change in time. From statistics and artificial intelligence with the application of database management. It is a rising area nowadays in good discipline. The idea of data mining is not to impose any rules on the datasets, but it helps us to guess and predict with a certainty analyzing only a very small set of data. Recently, data mining has gained a great attention in the knowledge and information industry due to the big data

2. Related Work

Many studies have been conducted on the performance of decision tree and on back propagation Classification is a main problem in

machine learning and data mining. Decision trees are well known because they are practical and in the user understandable format. Rules can also be extracted easily from a decision tree. Many algorithms such as J48, decision stump have been designed for constructing a decision tree. The usage of neural networks in classification is not new in machine learning community



3. WEKA interface

WEKA (Waikato environment for knowledge analysis) is a well known platform of machine software written in java, developed at the University Of Waikato New Zealand. The WEKA suite consists of a collection of visualization tools and various algorithms for analysis of data and predicting the future, along with graphical user interfaces for an easier access to this functionality. The actual non-Java version of WEKA was TCL/TK front-end software utilized to model algorithms implemented in other programming languages and data processing utilities in C. This Java version (WEKA 3) is used in various application areas such as educational purposes and research. There are many Advantages if WEKA:

1. It is a free source tool found in the World Wide Web
2. It is portable as it is completely implemented in the Java programming language and thus runs on any operating system.
3. It is vast collection of both the data processing and the modeling techniques.

It is simpler to use because of its graphical user interface.

4. WEKA supports many standard data mining tasks in particular data Preprocessing, clustering, classification, regression, visualization, and feature selection. All the above techniques are used to predict the future with one assumption that the data is available in only one single flat file without the violation of any redundancy principles and consistency principles.

4. Methodology

Weka allows .arff , .csv formats to perform data mining tasks. Here we are considering Attribute Relational File Format which is

@relation ebolavirus
@attribute fever {high_fever,normal_temperature}
@attribute body_temperature real
@attribute internal_bleeding {TRUE,FALSE}
@attribute weakness {TRUE,FALSE}
@attribute headache {TRUE,FALSE}
@attribute sore_throat {TRUE,FALSE}
@attribute trouble_breathing {TRUE,FALSE}
@attribute affected_days real
@attribute ebola_effected {yes,no}

Datasets:

high_fever,101,TRUE,TRUE,TRUE,TRUE,TRUE,4,yes
high_fever,104,TRUE,TRUE,TRUE,TRUE,TRUE,21,yes
normal_temperature,102,FALSE,FALSE,FALSE,FALSE,FALSE,3,no
normal_temperature,101,TRUE,TRUE,TRUE,TRUE,TRUE,7,no
high_fever,103,TRUE,TRUE,TRUE,TRUE,TRUE,8,yes
high_fever,105,TRUE,FALSE,TRUE,FALSE,TRUE,8,no
normal_temperature,98,TRUE,TRUE,TRUE,TRUE,TRUE,7,no
high_fever,104,TRUE,TRUE,TRUE,TRUE,TRUE,11,yes
high_fever,105,TRUE,TRUE,FALSE,FALSE,FALSE,3,no
high_fever,101,TRUE,FALSE,TRUE,FALSE,TRUE,9,no
high_fever, 100, TRUE, TRUE, FALSE, TRUE, TRUE, 9, yes
high_fever, 103, TRUE, FALSE, TRUE, FALSE,

TRUE, 8, no
high _fever, 102, TRUE, FALSE, TRUE, FALSE, TRUE, 8, no

5. Classification of J48 Algorithm

WEKA J48 algorithm is a famous machine learning algorithm depending on J.R. Quilan. The data or the data sets to be tested will be in the categorical format and therefore sequential data will not be tested at this stage. This will be again tested against C4.5 for verification and validation purpose. In the WEKA tool the implementation of an algorithm is always encapsulated in a class. And it also depends on other classes for their functionality. Every time the Java virtual machine executes J48 it generates an instance of the particular class by allocating memory for construction and storage of the decision tree classification. The algorithm, output procedure, classifiers, all these are the parts of the J48 algorithm.

Complex programs are to be always divided into the more than a class. While considering the J48 algorithm it does not contain any code in order to build a decision tree. Which normally always includes instances of other classes?

6. What is Ebola?

Ebola is a disease to humans which is mainly spread by the affected animals. This virus is transferred upon the contact with the blood of the infected animals. Mainly, vomiting, diarrhea, rash follow, along with the mal functionalities of liver and kidney occurs as symptoms which start after two or three weeks (21 days) after the virus has been entered into the body.

To confirm the virus, diagnosis is done on the blood samples for the viral antibodies, RNA, or the virus itself.

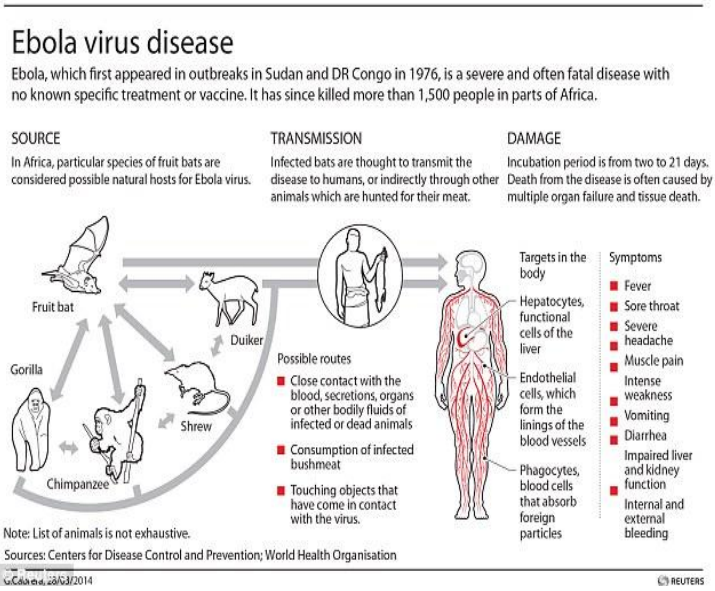
There is no specific treatment for this disease. Hence, one needs to take the precautions in order to be on the safe side of the EBOLA virus. The virus can also be transmitted by human-to-human contact or by the contact with the contaminated objects of the virus. The following are the signs and symptoms which indicate Ebola:

1. High fever
2. Headache
3. Joint and muscle pains
4. Sore throat
5. Weakness
6. Stomach ache
7. Lack of appetite

The following can significantly enhance the hopes for survival after the virus has been infected:

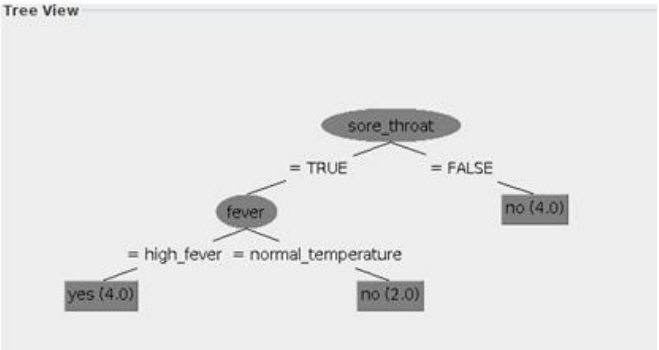
Provide intravenous fluids and balance the electrolytes.

Maintain the oxygen and the blood pressure domains periodically Treat all the associated infections.



7. Results

This below result from WEKA interface with J48 Machine Learning Algorithm will help to analyze a person symptoms indicating EBOLA have been analyzed in order to identify and cure the victims of Ebola virus



8. Conclusion

With the importance and increase of data in real world, the use of knowledge discovery and data mining techniques and tools have became important for analyzing the kinds of datasets in making decisions support. Here classification Techniques can be used to help analyze a person symptoms indicating EBOLA have been classified in order to identify and cure the victims of Ebola virus

References

1. Tina R. Patil, Mrs. S. S. Sherekar Sant Gadgebaba Amravati University, Amravati Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification
2. Xie, W. and Z. Huimin. A talent markets analysis method based on data mining. in Inteligent Computing and Inteligent Systems (ICIS), 2010 IEE International Conference on. 2010.
3. Kumar, V., et al., Asesing the Efect of Marketing Investments in a Busines Marketing Context. Marketing Science,201. 30(5): p. 924-940.
4. Aylmer, R.J, Who Makes Marketing Decisions in the Multinational Firm? Journal of Marketing, 1970. 34(4): p. 25-30.
5. Shaw, M.J, et al., Knowledge management and data mining for marketing. Decision suport systems, 201. 31(1): p.127-137.
6. Surma, J. and A. Furmanek. Improving Marketing Response by Data Mining in Social Network. in Advances in SocialNetworks Analysis and Mining (ASONAM), 2010 International Conference on. 2010.
7. Margaret H. Danham,S. Sridhar, ”Data mining, Introductory and Advanced Topics”, Person education , 1sted., 2006
8. Wenke Lee, Salvatore J. Stolfo, Kui W. Mok, “A Data Mining Framework for Building Intrusion Detection Models”