International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Post-Operative Patients Analysis using Data Mining Techniques

Anjana Menon, Swathi R, M Soumya Krishnan

Abstract: ***The post-operative patient analysis begins immediately after the surgery. The post-operative care depends on both the type of surgery and the previous health history. With the help of this analysis we can find if there exist any complications or not.***

In this paper we are analyzing the data of Post-Operative patients and we are identifying the state of the patient using J48 – a Data Mining Approach. After that, we are comparing various classification techniques in data mining and predicting their accuracy for post-operative patients data set.

We are comparing Naïve Bayesian, SMO, LWL, J48 classifiers using performance measures like Receiver Operating Characteristic), Root mean squared Error, Kappa statistics and Mean Absolute Error using WEKA tool. Different Techniques in Data Mining could be used for the extraction of data from large data sources. Data Mining Approaches can be used in several fields like Medicine, Education, Fraud Detection, Marketing etc.

Keywords: Data Mining, J48, Naïve Bayesian, WEKA

I. INTRODUCTION

Data mining is the extraction of patterns or knowledge from huge datasets. It is the method of evaluating patterns from different view and encloses it into advantageous knowledge for advance use. Knowledge discovery is used in various fields like education, business, banking etc. In medical field we use data mining to store and analyze large number of patient data and so they can optimize the efficiency and quality of their organizations. In this paper we are comparing four different data mining classification algorithms implemented on WEKA tool. By using the most accurate algorithm we are analyzing the post-operative patient dataset. The classifiers used are Naïve Bayesian, SMO, J48, LWL and the comparison depends on the performance measures like Receiver Operating Characteristic, Mean Absolute Error, Kappa statistics, and Root mean squared Error. The dataset contains nine attributes. They are L-CORE (internal temperature of the patient in Celsius), L-SURF (surface temperature of the patient in Celsius), L-O2 (Saturation of Oxygen in percentage), L-BP (last measurement of blood pressure),SURF-STBL (stability of surface temperature of the patient), CORE-STBL (stability of core temperature of the patient), BP-STBL (stability of blood pressure of the patient), COMFORT (the perceived comfort of the patient at discharge, measured as an integer in the range of 0-20), decision ADM-DECS (discharge decision).

Using these general details of post-operative patients we are analyzing the number of patients are in any one of the states mentioned below :-

I (patients sent to the Critical Care Unit)
S (patient suggested to go home)
A (patient sent to the ward of same hospital)

II. RELATED WORKS

Much research has been carried out using data mining in the field of medicine. In the research article[1], The author proposes a strategy for supporting doctors to analyze injury after surgery on the basis of an intelligent portable equipment that bind mechanized revelation of wounds as well as the recognition of contagious wounds. We incorporate pigment Standardization, vigorous peak disclosure, and derma shade discovery to disclosure injury. In the research article[2], The paper introduces various models of post - operative estimation of recurrence nomogramming the patients with renal tumor beyond meta states, and new entity depend over a multi-layer neurological chain are sketched to reduce and merge these method, that are thought into facilitate postsurgical monitoring protocols for surgeon. In research article[3], The article presents a microchip inserted patella corset developed for replicating the physical therapist's iris and training skill of getting systematic pneumatic patella calculations and associating the physiotherapist expertise to the wearer's assessment. In research article[4], Researchers discuss algorithms for the classification of data mining for prediction of kidney disease. The main objective is to predict and find the best classification algorithm using Naïve Bayes and Support Vector Machine based on accuracy and execution time. The output obtained after the experiment shows that the SVM has higher performance as comparing to Naïve Bayes. In the research article[5], In supervised data mining cases, researchers use two new approaches to synthesis classifiers. One is Rule-based which depends on the concealing of genomes while other is tree based classifiers that uses genetic programming to create decision tree. In the research article[6], the researchers compares different classification techniques using WEKA tool for emphasizing the mental condition about the speaker into either of the two classes- normal or aroused. The primary goal of the article is to identify the emotional state of the speaker and to compare the classifiers with the most accurate. In the research article[7], the researchers aim to predict if a person suffers

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Anjana Menon*, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.
Swathi R, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.
M Soumya Krishnan, Dept. of Computer Science and IT, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi, India.

from Parkinson's disease. The sound recording dataset of patients can be used for the prediction followed by using decision tree algorithm(CART).The voice of certain patients are taped and further transformed into voice attributes. These recordings are tested and identifies whether the patient is suffering from the disease and it also state the condition. In the research article[8], The researchers predict the burn patients ' sustainability. The algorithm of c4.5 is used to differentiate the dataset of sufferer applying WEKA. This algorithm's efficiency is examined using its accuracy, sensitivity, specificity and confusion matrix of classification. The results are found to be definite and correct by correlating with recorded information on sustainability or death. In research article[9], the researchers sum up various techniques for data mining, review and technical articles on diagnosis and prognosis of fibroid. Here, they bear to give an overview of the current research being conducted using data mining techniques to build up the diagnosis and prognosis of Fibroid. In the research article[10], the researchers defines the recognition lifestyle activities of user at home. The methods to describe Knowledge Discovery system for logging, sensing, mining, analyzing, recognizing and measuring the daily activities of user. In the research article[11], the researcher obtain effective system for the prognosis of cardiac disease using Naïve Bayesian. The person needs to cater deserved attribute values which is carried out as web based operation. Here two operations are performed classification and prediction. The accuracy obtained after using Naïve Bayesian is86%. In the research article[12], The researcher compared various classification technologies and precision in predicting ceaseless renal infection in general. Using WEKA tool Naive Bayes, J48, SVM, random forest and k-NN classification algorithms are compared. The output displays that random forest classifier has greater performance in classification than the other classifiers.

III. WORKING METHODOLOGY

In the analysis, we collected datasets of almost 51 postoperative patients. We gathered these data sets from several clinical sources. We analyze dataset using Naïve Bayesian Classifier in WEKA tool. As the result of these analysis we identified the patient's state and helps to decide whether patient must be taken to Critical Care Unit, patient sent to general ward of the hospital or permitted to leave the hospital.

A. *Datasets*

The sample datasets of post-operative patients are shown in the Fig.1.

B. *Mining Techniques*

Naïve Bayesian :- Bayesian approaches are an essential DM technique. Due to the distribution of probabilities, the Bayes classifier can demonstrably achieve the best result. The Bayesian method is based on the theory of probability. Bayes Rule is used to calculate from the last to the first, because the later two are generally easier to calculate from a probability model. Its noticeable that Naïve Bayes technique will not be a better option in certain scenarios.

Naïve Bayesian Classifiers :- Naive Bayesian technique is a supervised machine learning task. As it is easy and effective, this classification technique is considered to be important in data mining. Some assumptions have been made in naive bayes, and also it works quite well. The theorem states that,

$$P(A|B) = P(B|A)P(A) \div P(B)$$

Here,

P(A|B) : Conditional probability of occurrence of event A given event B istrue.
P(A) and P(B) :Probabilities of the occurrence of the events A and Brespectively.
P(B|A): probability of occurrence of the event B given event A istrue.

	%L-CORE (	L-SURF (pa	L-O2 (oxyg	L-BP (last	SURF-STBI	CORE-STB	BP-STBL	COMFORT	decision ADM-DECS (discharge decision)
2	mid	low	excellent	mid	stable	stable	stable	15	A
3	mid	high	excellent	high	stable	stable	stable	10	S
4	high	low	excellent	high	stable	stable	mod-stabl	10	A
5	mid	low	good	high	stable	unstable	mod-stabl	15	A
6	mid	mid	excellent	high	stable	stable	stable	10	A
7	high	low	good	mid	stable	stable	unstable	15	S
8	mid	low	excellent	high	stable	stable	mod-stabl	5	S
9	high	mid	excellent	mid	unstable	unstable	stable	10	S
10	mid	high	good	mid	stable	stable	stable	10	S
11	mid	low	excellent	mid	unstable	stable	mod-stabl	10	S
12	mid	mid	good	mid	stable	stable	stable	15	A
13	mid	low	good	high	stable	stable	mod-stabl	10	A
14	high	high	excellent	high	unstable	stable	unstable	15	A
15	mid	high	good	mid	unstable	stable	mod-stabl	10	A
16	mid	low	good	high	unstable	unstable	stable	15	S
17	high	high	excellent	high	unstable	stable	unstable	10	A
18	low	high	good	high	unstable	stable	mod-stabl	15	A
19	mid	low	good	high	unstable	stable	stable	10	A
20	mid	high	good	mid	unstable	stable	unstable	15	A

Fig.1 Screenshot of the Dataset

How does the Naive Bayes Algorithm work?

Step 1: Create a frequency table for all the features against the different classes.
Step 2: Draw the likelihood table for the features against the classes.
Step 3: Calculate the conditional probabilities for all the classes.
Step 4: Calculate max$iP(Ci|x1,x2,\ldots,xn)$

SMO :- Sequential minimal optimization (SMO) is a quadratic programming algorithm(QP) The drawback that comes throughout the guidance of support vector machines. SMO is an incremental heuristic used to solve the problem of computation. SMO fragments the mentioned drawback in an array of minimal achievable sustainable problems, that is rationally resolved. Results given by SMO classifier is as follows:-

- Obtain a wide range multiplier α1 that violates the optimization problem conditions of Karush - Kuhn - Tucker (KKT).
- Select a second α2 multiplier and optimize the com bination (α1, α2).
- Continue these steps until they are merged.

J48 :- J48 is a C4.5 extension that generates a decision tree using the C4.5 classifier. The decision tree generated can be used for classification and is therefore also known as statistical classification. The important part to remember when using the classifier is that the directory must be correctly structured and information's are correctly analyzed.
LWL :- Locally Weighted Learning is a lazy learning algorithm. To assign item weights, it uses an item-based algorithm. As a lazy learner, data processing is deferred until the results of a query are required.

Tool Used :- WEKA
Weka is data mining software using a collection of machine learning algorithms. These techniques are directly adapted to the dataset or by Java code. It consist different tools that can be used for :-

- Regression
- Clustering
- Association
- Data pre-processing
- Classification
- Visualization

Here, in this paper we used WEKA for Classification.
Classification :- We have classifiers in Weka to predict nominal or numerical quantities. Accessible system used for study are decision - making trees and lists, vector support machines, instance - based classifiers, logistic regression and networks of Bayes. All tabs are enabled once the data has been loaded. On the basis of requirements and by test and fault, the most appropriate algorithm can be found for an easily comprehensible representation of the data.
We need to set test options before running any classification algorithms. Below are available test options.
Used training sets:- Evaluation is based on how well the class of instances on which it has been trained can be predicted.

- Trained sets provided:-Assessment depends on the predictability of the class of instances loaded from a file.
- Cross-proofing: The assessment depends on cross - proofing of folds recorded in the text box "Folds."
- Split percent: Assessment depends on how good a certain percent of the data's can be predicted by using the values recorded in the ' percent ' field for testing.
- WEKA uses classifiers to categorize the dataset depended on the attribute aspects.

Data formats used in WEKA:- By default, Attribute Relation File format is used by WEKA to analyze the data's. However other formats supported by WEKA are listed below, from which data's can be carried:

- Comma-Separated Values
- Attribute-Relation File Format
- Database using Open Database Connectivity

C. *Result*

Result of this analysis will be the identified state of a post operative patient. The output is shown in Fig.2 and Fig.3 :-

Accuracy Measures :-

Algorithm	Kappa Statistic	Mean absolute error	Root mean squared error
Naïve Bayesian	-0.1198	0.2884	0.4079
SMO	0	0.2881	0.3741
J48	0	0.2791	0.3682
LWL	-0.064	0.2664	0.3782

Fig.2 Classification performance Algorithms based on error rate and kappa statistics

Performance of classifiers on the basis of classified instances :-

Classification Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)
Naïve Bayesian	18 (66.6667%)	9 (33.3333%)
SMO	20 (74.0741%)	7 (25.9259%)
J48	20 (74.0741%)	7 (25.9259%)
LWL	19 (70.3704%)	8 (29.6296%)

Fig.3 Performance based on classified cases

From the above results the more efficient and accurate data mining classifiers are :- SMO and J48.

So we are using J48 for analysis of the Post - Operative patient dataset to identify the current state of the patient. Confusion Matrix Obtained is shown in Fig.4 :-

```
a  b  c   <-- classified as
20 0  0   | a = A
1  0  0   | b = I
6  0  0   | c = S
```

Fig.4 Screenshot of confusion matrix

IV. CONCLUSION AND FUTURE WORK

The State of Post-Operative Patients are analyzed using the most accurate classifier among the four common classifier. The classifiers we have chosen are Naïve Bayesian, J48, LWL and SMO. We have found that J48 is the most accurate classifier and also depends on Receiver Operating Characteristic, Kappa Statistics , Root mean squared Error and Mean Absolute Error. Dataset analysis is implemented in WEKA tool. In future, the same dataset can be analyzed using the Hidden Markov Model(HMM).

REFERENCES

1. Yi-Chong Zeng, Kuan Hao Liao,Chu-Hsuan Wang,Yuling Lin,Wen-Tsung Chang, "Implementation of post-operative wound analytics", 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)
2. Baran Tander, Atilla Özmen, Ender Özden, " Neural network design for the recurrence prediction of post-operative non-metastatic kidney cancer patients",2015 9th International Conference on Electrical and Electronics Engineering (ELECO)
3. Nathan England, Samreen Islam, Henry Licht, Alexander Sarris, Caroline Twigg, "Sensor embedded knee sleeve for post-operative ACL patients", 2018 Systems and Information Engineering Design Symposium (SIEDS)
4. Guneet Kaur , Ajay Sharma " Predict chronic kidney disease using data mining algorithms in hadoop",2017 International Conference on Inventive Computing and Informatics (ICICI)
5. Thomas Weise,Raymond Chiong,"Evolutionary data mining approaches for rule-based and tree-based classifiers", 9th IEEE International Conference on Cognitive Informatics (ICCI'10)
6. T. Justin, R. Gajšek,V. Štruc, S. Dobrišek, "Comparison of different classification methods for emotion recognition", The 33rd International Convention MIPRO
7. S R Sonu, Vivek Prakash,Ravi Ranjan,K. Saritha "Prediction of Parkinson's disease using data mining", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)
8. M. Patil, Durga Toshniwal,R. C. Joshi, "Predicting Burn Patient Survivability Using Decision Tree In WEKA Environment", 2009 IEEE International Advance Computing Conference
9. K Girija, M. S. Shashidhara, "Data mining techniques used for uterus fibroid diagnosis and prognosis", 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)
10. Young-Sung Son, Topi Pulkkinen,Jun-Hee Park , "Active monitoring for lifestyle disease patient using data mining of home sensors",2013 IEEE International Conference on Consumer Electronics (ICCE)

AUTHORS PROFILE

Anjana Menon, PG Student, Master of Computer Applications, Amrita Vishwa Vidyapeetham School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi– 682 024, Kerala.

Swathi R, PG Student, Master of Computer Applications, Amrita Vishwa Vidyapeetham School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi– 682 024, Kerala..

M. Soumya Krishnan, working as Assistant Professor in CS & IT Dept., Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Kochi Campus, has over 16 years of experience in teaching. She has done MCA (B'Level) , M.Sc. CS and M.Phil. in Computer Science. She is currently doing her PhD in Computer Science. Her areas of interests are Data Analytics and Business Intelligence, Software Engineering, Data mining, Software Development Privacy and ITES.