International Journal of Engineering Research and General Science Volume 3, Issue 3, Part-2 , May-June, 2015
ISSN 2091-2730

Comparison of Six Classification Techniques for Post Operative Patient data in the Medicable discipline

Chinky Gera[1], Kirti Joshi[2]

Research Scholar[1], Assistant Professor[2]

Department of Computer Science & Engineering, RIMT-IET, Mandi Gobindgarh, India

chinkygera465.cg@gmail.com[1]

Abstract— Medical databases have accrued prodigious amount of enlightment regarding patients and their medical provision. The salient techniques of medical data mining incorporating post treatment of medicative data, fast, robust mining algorithms and reliability of mining results. The contemplate of this research paper is to provide a review on mining medical dataset, problem formulation and short description of preceding research in mining medical data. The experimental upshot concludes that surface temperature of the patient is not required and shows that while comparing decision table and J48 both has outperformed.

Keywords— Classification techniques; decision table; pre-processing; J48; accuracy; medicable; weka

INTRODUCTION

Data mining is the enactment of involuntary piercing the vast reserve of data to uncover the patterns and trends that progress afar manageable analysis. Data mining is usually to explain and understand the past behavior or to predict future behavior. Data mining plays an important role in healthcare by gathering and arranging the information about various reactions. Data Mining grasps the evident future for the healthcare field to empower the systems to systematically use data and analytics to pinpoint the inefficiencies and leading practices that amend care and rebate costs. Postoperative care is awareness of the potential reactions and complications of procedure. Postoperative care appears from the recovery room and continues throughout the recovery period. Patient's care must be carried out immediately postoperatively in the hospitals. While in a recovery, patient's body temperature, oxygen saturation, blood pressure, etc. will be monitored and when patient's condition seems to be stable moved to the hospital room where post operative care still continues or released to the home based on doctor's decision. The plan of this paper is to predict appropriate classification techniques that underpin the patient's care postoperatively with good accuracy.

The paper is sorted as underneath. In Section II, depicts the problem formulation. Some recent inspection of affiliated work in the medical field of data mining has been presented in Section III. Section IV scans the experiments and results accompanied by the Post Operative Patient dataset which is mined. Finally Section V discloses the conclusion with their future scope.

PROBLEM FORMULATION

There is a need to predict the best algorithm by comparing different classification algorithms. Pre-processing is very important to consider which contains useless attributes in the Post Operative patient dataset. It is necessary to remove those useless attributes to enhance the accuracy in this research paper. This can be run by enacting a respective research of classification algorithms containing 90 instances and 9 attributes including one class attribute. [2]

LITERATURE SURVEY IN THE MEDICAL FIELD

Diverse studies have been catalogued on pursing the machine learning action for the reliability analysis and predictive analysis. In this segment, abundant papers are assessed corresponding to data mining favourable in the medical sphere. The mining of data processes contains different techniques which are doctrinal in the healthcare field. The main aim of this paper is to study the data

mining techniques which are essential for medical data mining, normally to increase the accuracy of dataset.

A survey has been conducted on the current techniques in heart disease prediction by using knowledge discovery of data mining techniques. The mechanisms of predictive data mining on same dataset divulge decision tree outruns and sometime the Bayesian technique possess alike accuracy as of the decision tree and genetic algorithm is used further enhance the accuracy to attain optimal datasets which lessen the real data size helpful in forecasting heart diseases. [4]

Data mining methods and tools are used to produce the information from medical datasets associated with breast cancer disease to lessen time and effort and to aid the specialist to predict prior disease. The model retains 93.467% precise accuracy in testing set and the training set attains 96.8%. [5]

The number of blood donors and the blood group of particular age reveal the data mining model that use actual world data has been collected from EDP department of blood bank centre which uses the J48 algorithm for classification of the donors, help blood bank owner to gather proper decisions rapidly and more precisely. The results exhibit accuracy of 89.9% rate. [6]

Various data mining techniques are utilized to boost accuracy on the breast cancer diagnosis and prognosis. The results display that decision tree is formulated as better predictor having 93.62% of accuracy on the benchmark dataset and on the SEER dataset. [7]

A Hybrid approach is used of CART classifier with feature selection as well as bagging approach is used for analyzing the various datasets related to breast cancer. Training data is tested by using 10- fold cross validation. Bagging method is used to improve the decision tree Experiments performed with the combination of cart classifier, pre-processing and bagging used to enhance the classification accuracy of selected datasets. [1]

K-means method is used for dealing with clustering of medical database. In order to raise the efficiency of mining functions, some pre-processing methods used to seize 81% of accuracy and then by applying the algorithm again 94% accuracy was acquired for data amelioration. After that with current instances (700 records) yields 97% accuracy. [8]

K-Nearest Neighbour achieves higher accuracy of 97.4% in diagnosis of heart disease patients than neural network ensemble. By applying the voting to K-Nearest neighbour could not enhance the accuracy in diagnosis of patients suffering from heart disease. [12]

The comparison of different applications of data mining in healthcare region for releasing the useful information which parade 97.77% of accuracy for the prediction of cancer and success rate of the IVF treatment estimating around 70%. Developing the relevant data mining tools in terms of the human resources and skills lessen the cost and the time. [9]

The conversion of the raw data into useful information and to assist in detecting the patterns to determine the future trends in the medicable environment. The outcome reveals that decision trees are reliable and powerful decision making method imparts high accuracy of classification and helps the specialists to validate and organize the results to test and scan the various new symptoms of diseases which are based on the data. [10]

The main concentration was on the prediction of the unrevealed primary tumors in dataset where multiclass random forest classifier is used for the classification of the multiclass dataset gives higher accuracy as compared to the binary classifiers. For imbalanced dataset, SMOTE method is used for improvement of the results of selected classifier with greater accuracy. [3]

Different classification techniques have been compared equivalent to decision tree, Bayesian classification and concepts correlated with fuzzy. In first step, the outcome unveils that training dataset is better than use 10 cross fold. After that weka tool is utilized where ID3 algorithm is predicted as perfect for this work analysis. [11]

PROGNOSIS OF ALGORITHM: EMPIRICAL RESULTS AND EXPLORATION

This research paper is a formal commence seeks to apply inconsistent classification algorithms of data mining escorted by different statistics. Therefore, with the unveiling of enriched and the strained prediction techniques, there is need for an analyst to originate an algorithm which works best for a particular dataset.

Composed Dataset

For analysis of following data that are in the UCI repository are handed-down: Post-Operative Patient dataset gathered from the University Medical Centre, Institute of Oncology, Ljubljana and Yugoslavia [2]. Dataset include 90 instances having multi-variable data. Information of attributes of patients features are as follows.

- L-CORE: patient's core temperature in C.
- L-SURF: patient's surface temperature in C.
- L-O2: oxygen saturation in %
- L-BP: last measurement of blood pressure
- SURF-STBL: stability of patient's surface temperature
- CORE-STBL: stability of patient's core temperature
- BP-STBL: stability of patient's blood pressure
- COMFORT: patient's perceived comfort at discharge, measured as an integer between 0 and 20.
- Discharge ADM-DECS: discharge decision - Class I: patient sent to Intensive Care Unit,
  - Class S: patient prepared to go home,
  - Class A: patient sent to general hospital floor

Data Pre-processing

- Remove an attribute L-SURF by using unsupervised filter as shown in the Table I column: "After (Total 8 attributes)".
- Remove another attribute SURF-STBL also as shown in the Table I column: "After (Total 7 attributes)".

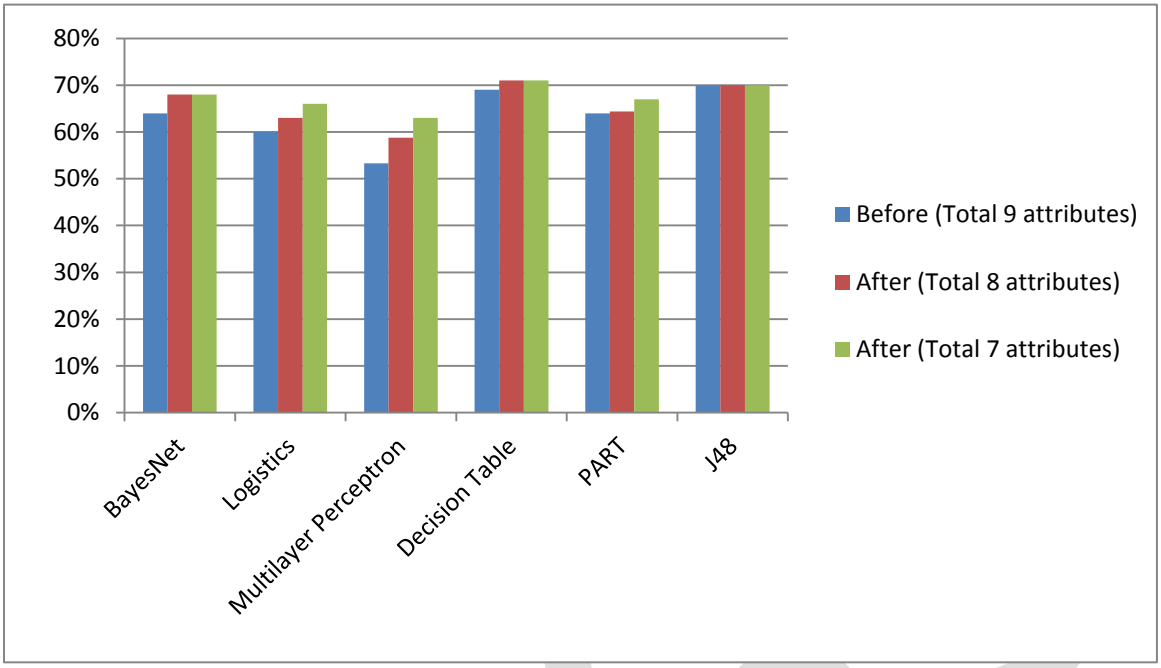
The surface temperature of the patient is not required and there is no need of more information, so we remove the useless attributes with this the accuracy can be enhanced.

Data Classification

For the enactment of machine learning algorithms, WEKA software is used. The experiments recounted in this research paper were accomplished using libraries from the Weka machine learning environment. Numerous classifiers are used in this paper. The following result shows that after removing the attributes L-SURF and SURF-STBL accuracy can be enhanced. In this paper, Table I shows that decision table achieves 71% of higher accuracy but J48 achieves 70% of same accuracy. It concludes that both decision table and J48 outperformed. Fig. 1 shows the graphical view of the comparison of six different classification algorithms.

ACCURACY COMPARISON OF SIX CLASSIFICATION TECHNIQUES

ALGORITHM	BEFORE (Total 9 attributes)	AFTER (Total 8 attributes)	AFTER (Total 7 attributes)
BayesNet	64%	68%	68%
Logistics	60%	63%	66%
Multilayer Perceptron	53.3%	58.8%	63%
Decision Table	69%	71%	71%
PART	64%	64.4%	67%
J48	70%	70%	70%

Graphical 

view of

Comparison of different Classification Techniques

ACKNOWLEDGMENT

Foremost, I would like to convey my sincere gratitude to my guide, parents and friends for their continuous support and immense knowledge. I would like to thanks them and wish them all the best in their lives.

CONCLUSION

In this paper the comparison has been performed to forecast the suitable classification algorithm for the post operative patient dataset which shows surface temperature of the patient is not required and it concludes from the experiment that decision tree and J48 outperformed to enhance the accuracy. The idea of the future work is to develop a new class and endorse in scheming medical decision support arrangement with the assist of selected algorithm.

REFERENCES

D. Lavanya and Dr. K. Usha Rani "Ensemble Decision Tree Classifier for Breast Cancer Data" International Journal of Information Technology Convergence and Services (IJITCS) Vol.2, No.1, February 2012.

Dataset collected, [http://tunedit.org/repo/UCI] accessed 2015.

Mehak Naib and Amit Chhabra "Predicting Primary Tumors using Multiclass Classifier Approach of Data Mining" International Journal of Computer Applications (0975 – 8887), Volume 96, No. 8, June 2014.

Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction" International Journal of Computer Applications (0975-8887), Vol. 17, No. 8, Mar. 2011.

Samar Al-Qarzaie, Sara Al-Odhaibi, Bedoor Al-Saeed and Dr. Mohammed Al-Hagery "Using the Data Mining Techniques for Breast Cancer Early Prediction".

Arvind Sharma and P. C. Gupta "Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool" International Journal of Communications and Computer Technologies, Volume 01, No. 6, ISSN Number: 2278-9723, September 2012.

Shweta Kharya "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 2, No. 2, April 2012.

Dr. Bushra M. Hussan "Data Mining based Prediction of Medical data using K-means algorithm" Basrah Journal of Science(A), Vol. 30(1), 46-56, 2012.

M. Durairaj and V. Ranjani "Data Mining Applications in Healthcare Sector: A Study" International Journal of Scientific and Technology Research, Vol. 2, ISSN 2277-8616, Issue 10, October 2013.

Aarti Sharma, Rahul Sharma, Vivek Kr. Sharma and Vishal Shrivatava "Application of Data Mining - A Survey Paper" International Journal of Computer Science and Information Technologies, Vol. 5(2), 2023 – 2025, 2014.

Satya Ranjan Dash and Satchidananda Dehuri "Comparative Study of Different Classification Techniques for Post Operative Patient Dataset" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 5, July 2013.

Mai Shouman, Tim Turner and Rob Stocker "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients" International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012