Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Classifier Approach of Life Prediction using THORACIC Dataset.

D Narasimhan[1]*, S Bhuvaneswari [2], R Bala Krishnan[3], and A Srinivasan[4].

Department of Mathematics[1], Department of CSE[3], Department of ECE[4] SASTRA University[1,3,4], Srinivasa Ramanujan Centre, Kumbakonam, ARJ College of Engineering[2], Mannargudi, Tamilnadu, India

ABSTRACT

Knowledge discovery from data in data mining has been a great success in many applications such as business, marketing, banking, transportation, medicine and health care. Data mining enables characterization of the problems in many surgeries. Data mining is used for decision making for surgery and provides data on which combinations have failed and helps in minimizing the failure rate. This paper intends to predict the life survey for post operation of thoracic surgery. The data set considered for the present study is taken from UCI machine learning respiratory, an online repository of large data sets. The obtained results from the tests help for Human life prediction and the intention is to classify need for Thoracic surgery.

Key words: Thoracic surgery, Training and Testing data, Naive Bayesian classifier

Corresponding author

INTRODUCTION

Surgery performed on organs and tissues in our midsection (Thorax) is called Thoracic surgery .It is a surgical strength which is cantered around systems including the midsection, all the more formally known as the thorax. This surgical claim to fame can be profoundly focused and greatly difficult. In the medicinal group, thoracic surgery has a notoriety for being a high-chance strength with a great deal of issues [4]. Thoracic specialists could treat lung malignancy, emphysema and different heart sicknesses [5]. The American numerical society assesses in 2014 there are 2,24,000 instances of lung tumor and more than 1,59,000 patient passing from lung growth .The fundamental driver of lung disease is smoking and breathing the contaminated air and so on [1]. Patients are alluded to a thoracic specialist when it gets to be evident that they have midsection conditions which require surgical treatment. Numerous complexities that happen from thoracic operations can be foreseen. This sort of more entangled surgery is finished by a few clinics and specialists while others done genuinely routine surgery. With the goal that they can make reasonable examinations between them, the death rates have been danger acclimated to consider the trouble of every operation [2]. The Society of Thoracic Surgeons is a non-benefit association speaking to more than 6,700 specialists, scientists, and unified wellbeing experts overall why should committed guaranteeing the best conceivable heart, lung and other surgical techniques for the midsection. The majority of the examination would help to break down future events.

MATERIALS AND METHODS

The information set for the infection is procured from UCI, an online archive of huge information sets. The Naive Bayesian (NB) grouping is utilized to characterize this thoracic surgery information [3]. Gullible Bayesian grouping is taking into account Bayes hypothesis and the classifier is a measurable classifier. The procedure of utilizing measurable method to find relationship between information things, and the development of prescient models in view of them. Naive Bayesian Classifier (NBC) expect that the impact of a trait esteem on a given class is free of the estimations of different traits [6-8]. This suspicion is called class restrictive freedom. This thoracic information have both ostensible and twofold qualities and the information sets are haphazardly parceled into preparing set and test information set. The information are at initially portioned by the class and afterward the back likelihood can be ascertained and building a recurrence table for every trait. At long last the NBC comparison can be connected to compute the back for every class. The classes with most noteworthy probabilities are the results of the forecast [9]. Grouping frameworks can help in expanding exactness and dependability of the judgment and minimize the conceivable slips, and additionally setting aside a few minutes productive. Guileless Bayesian procedure offers high exactness and velocity when connected to huge information base. Contrasting with all different classifiers, the Naive Bayesian classifier has the base slip rate and it holds the accompanying focal points:

- Fast, highly scalable model building and scoring.
- Works with both binary and multiclass classification.
- Needs only a small amount of training data set for acquiring knowledge.

Finally there exist some combinations of predictive result that will help us to find which case suitable for surgery in order to be more accurate .From the observations, the recommended cases are to be considered for surgery.

Thoracic Parameters (attributes):

Diagnosis, FVC, FEV1, Zubrod scale, Pain before surgery, Hemoptysis, Dyspnoea, Cough before surgery, Weakness before surgery, Size of the original tumor, Diabetes mellitus, MI up to 6 months ,peripheral arterial diseases, Smoking, Asthma, Age at surgery, Risk 1 year survival period.

Prediction Algorithm for Need for the Thoracic Surgery using Naive Bayesian Classifier
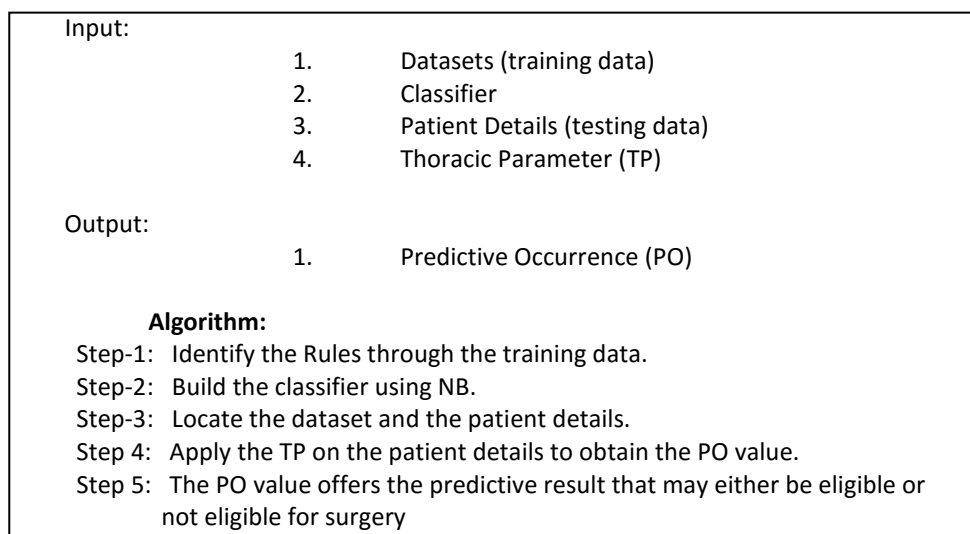
```
Input:
                1.      Datasets (training data)
                2.      Classifier
                3.      Patient Details (testing data)
                4.      Thoracic Parameter (TP)

Output:
                1.      Predictive Occurrence (PO)

        Algorithm:
  Step-1:  Identify the Rules through the training data.
  Step-2:  Build the classifier using NB.
  Step-3:  Locate the dataset and the patient details.
  Step 4:  Apply the TP on the patient details to obtain the PO value.
  Step 5:  The PO value offers the predictive result that may either be eligible or
            not eligible for surgery
```

Fig. 1: Procedure: 'Need for Surgery Prediction using Thoracic Data'

The proposed scheme for the need of surgery prediction has been tested using the dataset from UCI repository.

RESULTS AND DISCUSSION

The training dataset of UCI repository have been taken to build and train the naive Bayesian classifier and the classifier works with the testing dataset (patient details) using the knowledge from the trained data. The experiment was performed through the above mentioned procedure 'Need for Surgery Prediction using Thoracic Data' and the outcomes of training set building procedure have been stated in Table 1. The proposed scheme attempts to maximize the posterior probability in determining the class label that is either surgery needed or not. It predicts 1 year risk = False for each result. The test cases from the observation reveals that the people who are at high risk and through the outcomes one can estimate the issues in understanding the difficulties of a disease and improve the discriminating power of the rule.

Table 1: Success rate of Thoracic surgery needs prediction on UCI repository

FVC	Pain	Haemoptysis	Cough	Weak	Size	Smoke	Age	Probability (1 Year Risk)	
								False	**True**
< 3	F	F	T	T	OC11	T	> 50	0.0101298	0.0020805
< 3	F	F	T	F	OC11	T	≤50	0.0025508	0.0003612
< 3	F	F	T	F	OC11	T	>50	0.0537502	0.0069846
< 3	F	F	T	F	OC12	T	>50	0.0725201	0.0148424
≥3	F	F	T	T	OC11	T	≤50	0.0006619	0.0001187
≥3	F	F	T	F	OC11	T	>50	0.0740099	0.0077072
≥3	F	F	T	F	OC12	T	≤50	0.0047388	0.0008471
≥3	F	F	T	F	OC12	T	>50	0.0998546	0.0163778

The testing dataset i.e the patient details are subjected to the above classification process to obtain the data value for the class label 'Risk'. The results of prediction are shown in Table 2 and it shows that the prediction result "False" indicates the risk factor in the cases of persons with the habit of Smoking suffering from cough and weakness.

Table 2. Predictive outcome of the UCI repository testing dataset

FVC	Pain	Haemoptysis	Cough	Weak	Size	Smoke	Age	Risk
< 3	F	F	T	T	OC11	T	> 50	F
< 3	F	F	T	F	OC11	T	≤50	F
< 3	F	F	T	F	OC11	T	>50	F
< 3	F	F	T	F	OC12	T	>50	F
≥3	F	F	T	T	OC11	T	≤50	F

CONCLUSION

This implementation of the proposed classification and prediction measure for the UCI repository dataset using naive Bayesian considers 17 attributes with various ranges of values for the experiments. After thorough verifications and analysis, the attributes are reduced to 8 by neglecting the attributes which are not relevant or of less impact for thoracic surgery. Further Naive Bayesian technique was applied for life prediction to the limited (7) attributes and the extracted patterns were interesting and the results are verified by the validation. The evaluation pattern might be useful for analysing different types of healthcare data for life prediction. The paper addresses the effectiveness of the application of Bayesian classifier for prediction using the Thoracic Dataset.

ACKNOWLEDGEMENT

REFERENCES

[1]   Dr. R. Geetha Ramani, G. Sivagam, Int J of Comp Appl 2011; 32(9).
[2]   Hashmi, Sumaiya F., A Machine Learning Approach to Diagnosis of Parkinson's Disease (2013).
[3]   Shantakumar B.Patil, Y.S.Kumaraswamy, Eur J of Sci Res 2009; 31(4):642-656.
[4]   UCI Machine Learning Repository - Center for Machine Learning and Intelligent System.
[5]   K.Srinivas, B.Kavihta Rani , A.Govrdhan , (IJCSE) Int J on Comp Sci and Engg 2010; 02(02):250-255.
[6]   Indrajit Mandal, N. Sairam, Int J of sys sci 2012.
[7]   Carloz Ordonez, IEEE Tran on Inf Tech in Biomedicine 2006;10(2):334 - 343.
[8]   Harleen Kaur , Siri Krishan Wasan and Vasudha Bhatnagar, Data Science J 2006;5:119-126.
[9]   S.Lekshmi, G.P.Mohanta, K.G.Revikumar, P.K.Mannai, Int J of Phar and Pharmaceutical Sci, 2014;6(2):92-98.