

Analysis of Health care Data using Data Mining Techniques

Mahesh S Nayak¹, Dr. S.Kavitha², Dr. M. Hanumanthappa³

¹Research and Development Centre, Bharathiar University, Coimbatore – 641046

²Assistant professor, Department of computer applications, Dayananda Sagar Institutions,

³Professor, Department of Computer Science & Applications, Bangalore University, Bangalore.

Abstract - In the current generation, diabetes is a common and chronic disease found in eighty percent of the people in India. In this regards, very sparse literature is there in study on diabetes in computer science field. In our proposed method, the study gives cost benefit analysis of the diabetes dataset of 768 instances. In this research paper, data mining algorithms are used for classification and recorded the error rates and time complexity of all the algorithms. It is proved that the ZeroR algorithm given in the short period of time taken only 0.01 seconds. For any further research on the same health care data analysis, ZeroR algorithm can be used for the faster outputs.

I. INTRODUCTION

Data mining nothing but knowledge discovery in data that the practice of dynamically finding the patterns from big data. The knowledge discovery applies mathematical algorithms for evaluation and to make subgroups of data. The application of data mining is producing the meaningful information [6], big data analytics, tuned dataset or database to data mining tools. The risk factor for diabetes is the current research topic for researchers[1].



Figure 1: Reasons for Diabetes

II. LITERATURE SURVEY

Noman Sohail e,al proved that k-means given 96% accuracy of classification[2]. The common symptoms of diabetes discussed in the research[3]. In the proposed research[4] found that in United States 30 million people are suffering from diabetes. Medical specialists can determine diabetes by FPG [1]. The objective of this research is to find the data mining algorithm to present the classification with accuracy.

III. DATA SET

The dataset is used for the proposed method is Pima dataset which is sponsored by NID [5]. It is having 768 instances and 8 classes.

IV. EXPERIMENTS AND RESULTS

Preprocessing is carried out for the dataset and the attributes are shown in the Figure 2. The algorithms used for classification is ZeroR, PART, OneR and JRIP classification algorithm. Figure 3 shows the cost analysis and classification ratio of positive and negative of healthcare dataset tested using ZeroR algorithm. Figure 4 shows the PART algorithm classification analytics and Figure 5 shows the JRIP algorithm for classification [2].

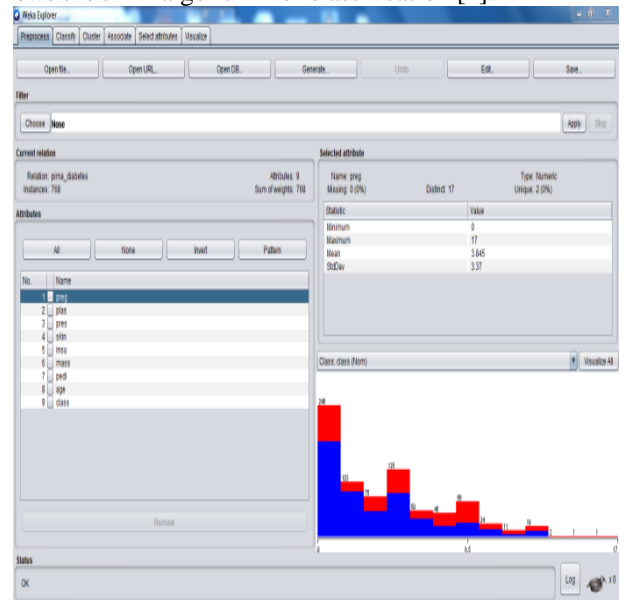


Figure 2: Preprocessing of dataset

V. CLASSIFICATION

Table1: Comparison of classification models

Method	ZeroR	PART	OneR	JRIP	BestFirst
Correctly Classified Instances	65.1042 %	75.260 4 %	71.48 44 %	76.04 17 %	71.224 %
Incorrectly Classified Instances	34.8958 %	24.739 6 %	28.51 56 %	23.95 83 %	28.776 %
Kappa statistic	0	0.439	0.322 6	0.453 8	0.3492

Mean absolute error	0.4545	0.3101	0.2852	0.3419	0.3448
Root mean squared error	0.4766	0.4149	0.534	0.4239	0.4277
Relative absolute error	100%	68.224 %	62.73 98 %	75.23 22 %	75.8525 %
Root relative squared error	100%	87.041 8 %	112.0 334 %	88.93 3 %	89.7294 %
Time taken	0.01 Sec	0.03Sec	0.02 Sec	0.05sec	0.09sec

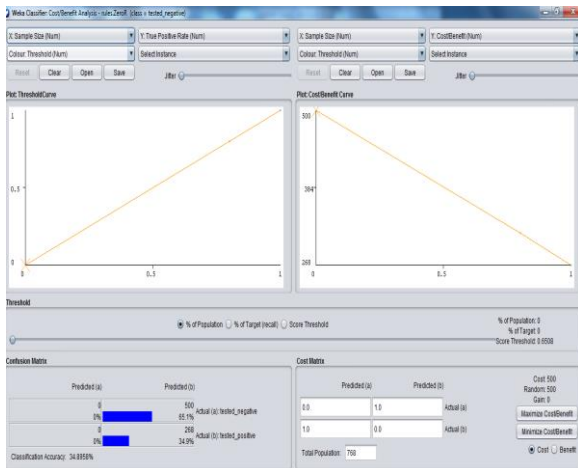


Figure 3: Cost estimation graph of Zero R method

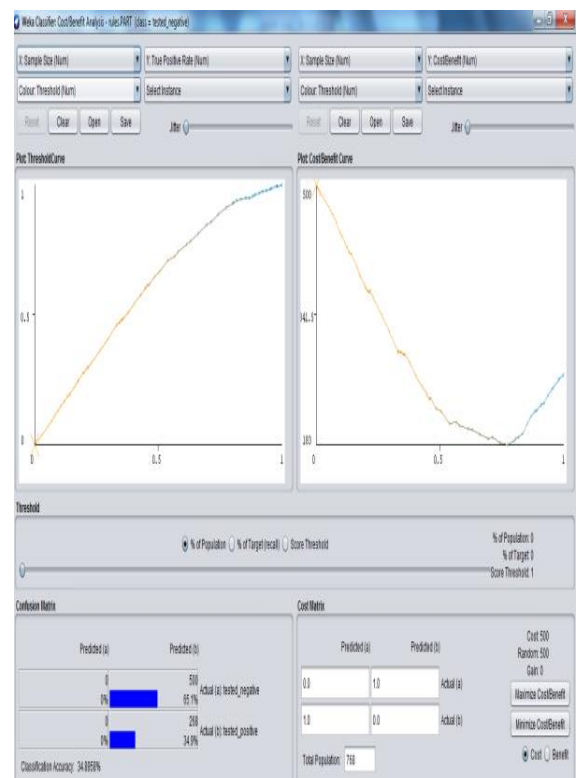


Figure 4: Cost estimation graph of PART method

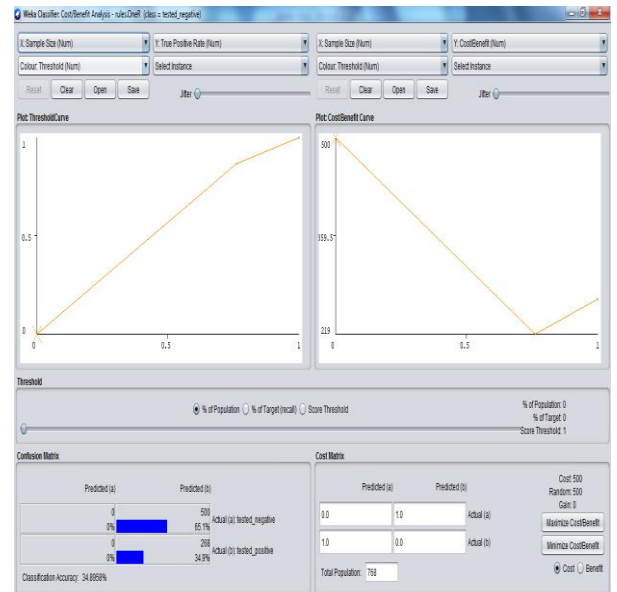


Figure 5: Cost estimation graph of One R method

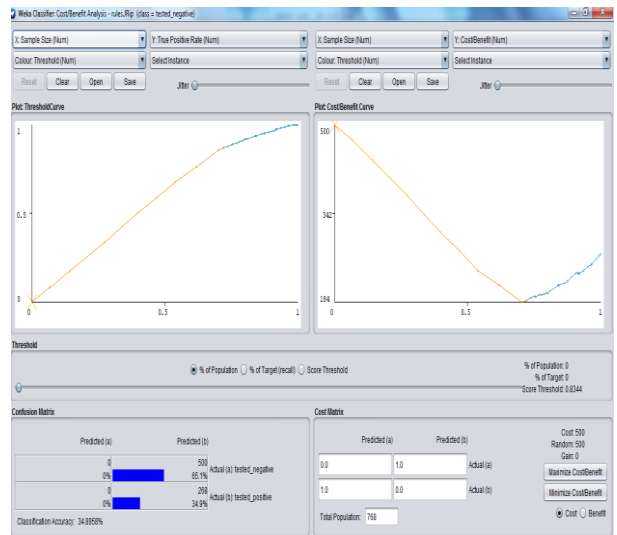


Figure 6: Cost estimation graph of JRip method

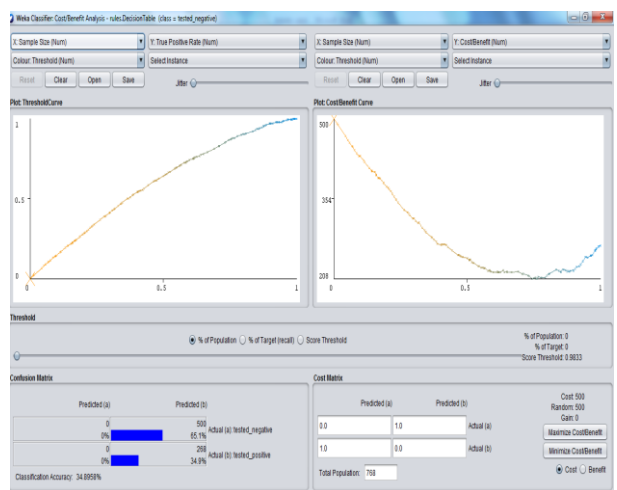


Figure 6: Cost estimation graph of Decision Table method

Decision Table:

Stale search after 5 node expansions
 Total number of subsets evaluated: 43
 Merit of best subset found: 77.604
 Evaluation (for feature selection): CV (leave one out)
 Feature set: 1,2,4,9

VI. CONCLUSION

The aim of our research paper is to analyze the health care dataset for choosing the best data mining algorithm. In the proposed method we have used ZeroR algorithm, PART algorithm, OneR algorithm, JRIP algorithm for classification as well as cost analysis. Our study proved that the ZoroR algorithm classified the 768 instances within 0.01 seconds. In the future work, we can use other health care dataset for the data using other data mining tools.

VII. REFERENCES

- [1]. Irshad Muhammad et al. An Indoor LOS & NON-LOS Propagation analysis using Visible Light Communication. Int. J. Adv. Res. Comput. Eng. Technol.7, 727–732 (2018).
- [2]. Noman Sohail et al., "classification and cost benefit Analysis of Diabetes mellitus" IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.10, October 2018
- [3]. Wolfsdorf JI, Glaser N, Agus M, Fritsch M, Hanas R, Rewers A, et al. Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State: A Consensus Statement from the International Society for Pediatric and Adolescent Diabetes. *Pediatr Diabetes* [Internet]. 2018 Jun;0(ja). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/pedi.12701>.
- [4]. Disease Control C. N. Diabetes Statistics Report Estimates of Diabetes and Its Burden in the US Background. 2017. Sohail MN, Jiadong R, Uba MM, Irshad M. A Comprehensive Looks at Data Mining Techniques Contributing to Medical Data Growth: A Survey of Researcher Reviews. In Springer, Singapore; 2019 [cited
- [5]. Pima Indians Diabetes Database | Kaggle.
- [6]. Dr.M.Hanumanthappa Dr.S.Kavitha, "Predictive Analysis of Different Classification Techniques of I Virus - h4N2 influenza using Weka", International Journal for Research in Applied Science and Engineering Technology, Vol.5, Issue IX, IJRASET(Google scholar)
- [7]. Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *Int J Med Sci.* 2014;11(11):1185–200.



Dr.Kavitha is completed her M.Phil and Ph.D in Computer Science. She is published many research papers in national and international journals She is the winner of International award "Outstanding Faculty in Science" and "Best teacher of the country award. Her area of research interest is Image rocessing,Data Mining,Neural Networks.



Mahesh S Nayak completed M.Phil in 2011. He is published many research papers in national and international journals.



Dr.M.Hanumanthappa is working as a professor in Bangalore University. He is published more than 100 research papers in national and international journal.He is chairman of BOS for Bangalore University.His area of research is Data Mining, Text processing