The efficiency of Hidden Naïve Bayes Technique compared with Data Mining Techniques in early diagnosis of diabetes and prediction system

Bassam Abdo Al-Hameli 1,2*, AbdulRahman A. Alsewari 1

1 IBM Centre of Excellence, Faculty of Computing, Universiti Malaysia Pahang,
26300, Kuantan, Malaysia,
2 Computer and Information Technology College, University of Sheba, Sana'a, Yemen.
* bassam566606@gmail.com

Abstract:
The efficiency of classification techniques largely varies on the techniques used and the data sets. A classifier process efficiency lies in how accurately, it categorizes the item. The technique of classification finds the relationships between the predictor's worth and the goal values. This paper is an in-depth study of the Hidden Naïve Bayes (HNB) classification technique compared to state-of-the-art techniques in the medical field, which have demonstrated HNB efficiency and ability to increase the accuracy of prediction. This study examines the efficiency of the four machine learning techniques including HNB, Decision Tree C4.5, Naive Bayes (NB), and Support Vector Machine (SVM) on the diabetes data set to identify the possibility of creating predictive models with real impact. The four classification techniques are studied and analyzed, then their efficiency is evaluated for the PID dataset in terms of accuracy, precision, F-measure, and recall, in addition to other performance measures. The outcome of this analysis shows that HNB is more reliable than other techniques.

Keywords: Classification; Hiden Naïve Bayes; Naïve Bayes;Decision Tree; Support Vactor Machine; Data Mining; Pima Indian Diabetes Dataset.

REFERENCES

[1] Jantawan, B. and C.-F. Tsai, The application of data mining to build classification model for predicting graduate employment. arXiv preprint arXiv:1312.7123, 2013.

[2] Foroughi, F. and P. Luksch, Data Science Methodology for Cybersecurity Projects. arXiv preprint arXiv:1803.04219, 2018.

[3]Rupa Bagdi, Prof. Pramod Patil," Diagnosis of Diabetes Using OLAP and Data Mining Integration", International Journal of Computer Science & Communication Networks,Vol 2(3), pp. 314-322.

[4] Shankaracharya, D.O., S. Samanta, and A.S. Vidyarthi, Computational intelligence in early diabetes diagnosis: a review. The review of diabetic studies: RDS, 2010. 7(4): p. 252.

[5] Palivela Hemant, Thotadara pushpavathi, "A novel approach to predict by cascading clustering and classification", Proc of the 3rd international conference on computing communication &networking technologies, Vol.5, No.3, pp.135-155, Dec 2012.