

The Accuracy of KNN, decision tree, random forest, SVM, neural network and naive Bayes classifier for early Prediction of Diabetes

Hsieh Cheng-Han

October 2022

abstract

This work compares the accuracy of some classifiers for early the prediction of diabetes. More specifically, the research compares the accuracy of k-nearest neighbors (KNN) classifier, decision tree, random forest, support vector machine (SVM), neural network, and naive Bayes classifier on the prediction of diabetes, which the dataset is collected with eight features, times of pregnancy, concentration of glucose in blood, blood pressure, skin thickness, concentration of insulin in blood, body mass index (BMI), the value of diabetes pedigree function and age.

The result show that XXX is the most accurate on the prediction of diabetes.

1 Introduction

Diabetes is a chronic disease which may cause many complications. There're lots of reason that can put a person at the highly risk of having diabetes, such as age, obesity, lack of exercises, and more on. So many reasons interweave together making the manual prediction on diabetes is nearly impossible. However, lots of works [1] [2] [3] show that it is possible to have high accuracy by using machine learning techniques, such as random forest, K-means clustering, neural network, and so on.

By collecting the essential data of human body, prediction of diabetes can be turn into classification problem. Imagine that an individual case with essential data is a point in hyperspace, if it is closer to the cluster having diabetes, this case is more likely to have diabetes in the future, otherwise, this case is more likely healthy. But there are lots of machine learning techniques born to solve clas-

sification problem, it remains a problem that which technique having the highest accuracy on the prediction of diabetes.

To find out which techniques is more suitable to predict diabetes, this work examines the diagnosis of diabetes using KNN classifier, decision tree, random forest, SVM, neural network and naive Bayes classifier.

2 Related works

3 KNN

4 SVM

5 neural network

6 Naive Bayes classifier

References

- [1] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 292–299, 2019, 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050920300557>
- [2] T. Mahboob Alam, M. A. Iqbal, Y. Ali, A. Wahab, S. Ijaz, T. Imtiaz Baig, A. Hussain, M. A. Malik, M. M. Raza, S. Ibrar, and Z. Abbas, "A model

for early prediction of diabetes,” *Informatics in Medicine Unlocked*, vol. 16, p. 100204, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914819300176>

- [3] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, “Predicting diabetes mellitus with machine learning techniques,” *Frontiers in Genetics*, vol. 9, 2018. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00515>