

Prediction of diabetes mellitus in bogotá using machine learning tools

Robert Sebastian Castellanos Rodriguez

National University of Colombia, Colombia

Abstract

Due to its continuously increasing occurrence, more and more families are influenced by diabetes mellitus. Most diabetics know little about their health quality or the risk factors they face prior to diagnosis. In this study, we have proposed a novel model based on data mining techniques for predicting type 2 diabetes mellitus (T2DM). The main problems that we are trying to solve are to improve the accuracy of the prediction model, and to make the model adaptive to more than one dataset. Based on a series of preprocessing procedures, the model is comprised of two parts, the improved K-means algorithm and the logistic regression algorithm. The colombian Diabetes Dataset and the Waikato Environment for Knowledge Analysis toolkit were utilized to compare our results with the results from other researchers. [1].

Keywords

K-means, Health prediction, Machine learning.

References

- [1] Castellanos, R.S. (2021). Using machine learning for predictions of disease.