Comparing and Tuning Machine Learning Algorithms to Predict Type 2 Diabetes Mellitus

Gabriel Aguilera-Venegas, Amador López-Molina, José Luis Galán-García
University of Málaga
gabri@ctima.uma.es, ama2lm@outlook.com, jlgalan@uma.es

Gemma Rojo-Martínez
IBIMA, Málaga, Spain
gemma.rojo.m@gmail.com

Abstract

The main goals of this work is to study and compare machine learning algorithms to predict the development of type 2 diabetes mellitus.

Four classification algorithms have been considered, studying and comparing the accuracy of each one to predict the incidence of type 2 diabetes mellitus seven years in advance. Specifically, the techniques studied are: Decision Tree, Random Forest, kNN (k-Nearest Neighbors) and Neural Networks.

The study not only involves the comparison among these techniques, but also, the tuning of the meta-parameters in each algorithm.

The algorithms have been implemented using the language R.

The data base used is obtained from the nation-wide cohort di@bet.es study [1].

The conclusions will include the accuracy of each algorithm and therefore the best technique for this problem. The best meta-parameters for each algorithm will be also provided.

References

1. ROJO-MARTÍNEZ AND G. AND VALDÉS AND S. AND SORIGUER AND F. ET AL.. Incidence of diabetes mellitus in Spain as results of the nation-wide cohort di@bet.es study.. Sci Rep 10, 2765 (2020). https://doi.org/10.1038/s41598-020-59643-7.