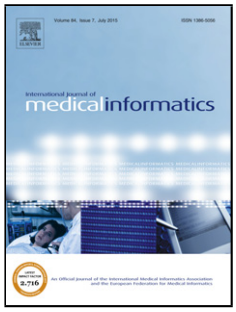
Accepted Manuscript



Title: Identifying People at Risk of Developing Type 2 Diabetes: A Comparison of Predictive Analytics Techniques and Predictor Variables

Authors: Amir Talaei-Khoei, James M. Wilson

PII: S1386-5056(18)30399-X
DOI: https://doi.org/10.1016/j.ijmedinf.2018.08.008
Reference: IJB 3734

To appear in: *International Journal of Medical Informatics*

Received date: 24-1-2018
Revised date: 26-7-2018
Accepted date: 16-8-2018

Please cite this article as: Talaei-Khoei A, Wilson JM, Identifying People at Risk of Developing Type 2 Diabetes: A Comparison of Predictive Analytics Techniques and Predictor Variables, *International Journal of Medical Informatics* (2018), https://doi.org/10.1016/j.ijmedinf.2018.08.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Identifying People at Risk of Developing Type 2 Diabetes: A Comparison of Predictive Analytics Techniques and Predictor Variables

Amir Talaei-Khoei

Department of Information Systems, Ansari College of Business, University of Nevada, Reno, USA
School of Software, Faculty of Engeering and IT, University of Technology Sydney, Australia
atalaeikhoei@unr.edu

James M Wilson

Nevada Medical Intelligence Center, School of Community Health Sciences and Department of Pediatrics, University of Nevada Reno , USA
jamesmwilson@unr.edu

Corresponding Author:
Dr. Amir Talaei-Khoei
University of Nevada Reno
atalaeikhoei@unr.edu
+1 775 440 7005

Highlights

1. The study aims to identify the patients at risk of type 2 diabetes (T2D).
2. The study compares the performance of machine learning classification algorithms.
3. A list of predictors for T2D for short, medium and long term is provided.
4. Period and Purpose of prediction of T2D influence the performance.

Abstract

Background: The present study aims to identify the patients at risk of type 2 diabetes (T2D). There is a body of literature that uses machine learning classification algorithms to predict development of T2D among patients. The current study compares the performance of these classification algorithms to identify patients who are at risk of developing T2D in short, medium and long terms. In addition, the list of predictor variables important for prediction for T2D progression is provided.

Methods: This study uses 10,911 records generated in 36 clinics from the 15th of November 2008 – 15th of November 2016. Syntactic minority oversampling and random under sampling were used to create a balanced dataset. The performance of Neural Networks, Support Vector Machines, Decision Tress and Logistic Regression to identify patients developing T2D in short, medium and long terms was compared. The measures were Area Under Curve, Sensitivity, Specificity, Matthew correlation coefficient and Mean Calibration Error. Through importance analysis and information fusion techniques the predictors of developing T2D were identified for short, medium and long-term risk analysis.

Results: The findings show that the performance of analytics techniques depends on both period and purpose of prediction whether the prediction is to identify people who will not develop T2D or to determine at risk patients. Oversampling as opposed to under sampling improved performance. 16 predictors and their importance to determine patients at risk of T2D in short, medium and long terms were identified.

Download English Version:

https://daneshyari.com/en/article/10132717

Download Persian Version:

https://daneshyari.com/article/10132717

Daneshyari.com