

Diabetes Patients Early Readmission Prediction

Peter Mačinec and František Šefčík

Faculty of Informatics and Information Technologies,
Slovak University of Technology, Bratislava

Abstract. Nowadays, more and more patients suffer from still incurable diabetes disease. Every wrong chosen treatment for patients can harm their health and lead to early readmission that costs more money. Therefore, there is a demand for predicting the readmission of patients to increase quality of health care and also to reduce costs. However, standard methods for identifying the patients with risk of readmission perform poorly (e.g. LACE index). With growing number of patients with diabetes, there is a need for methods that can automatically and more accurately predict the readmission. In this paper, we provide method based on machine learning for predicting early readmission of patient. The results of data analysis already showed that there is a potential of using data-driven approach for this problem.

Keywords: diabetes · early readmission prediction · machine learning · data analysis.

1 Motivation

Diabetes is a wide spread chronic disease that is related to irregular blood glucose levels caused by problems with insulin. The number of people with diabetes has increased enormously in recent years and costs for health care with each hospital admission are rising simultaneously. The methods of diabetes treatment of patient have high impact on mortality and morbidity. Wrong treatment can endanger patient's health and may lead to early readmission.

A hospital readmission is when a patient who is discharged from the hospital, gets re-admitted again within certain period of time. By predicting readmission, more attention may be given to treatment of patients with high probability of readmission and so increase the quality of care during hospitalization. Because there is no cure for the diabetes yet¹ and diabetic patients can be readmitted in the future, an *early readmission prediction* can help mostly when it comes to selection of best treatment for the patient.

Data of patients clinical encounters are being collected naturally with health-care systems, thus data-driven approach seem to be appropriate for this problem. Machine learning algorithms used for early readmission prediction provide ability to process the data of a lot of patients and may help to find hidden dependencies in the data to outperform basic methods (e.g. LACE index).

¹ <https://www.diabetes.org.uk/diabetes-the-basics/is-there-a-cure>

The task of *early readmission prediction* can be represented as binary classification problem into two classes - patient was early readmitted or not. The term *early readmission* is very relative. In our case, we define it as patient being readmitted in less than one month (30 days).

2 Dataset

Our task to predict diabetes patients early readmission was evaluated in offline experiments, performed on real world dataset of diabetes patients clinical encounters [1]. The dataset was created from Health Facts database with data collected during 10 years (1999-2008) across 130 hospitals in United States. From this large-scale database with millions of records, final dataset with 101 766 records of patients encounters was derived using 5 criteria defined by authors (e.g. it is diabetic and inpatient encounter, or that laboratory tests were performed during the encounter, etc.).

The dataset contains 50 attributes (features), both numerical and categorical. Attributes are of various types - demographics of the patient (like *race*, *gender*, etc.), diagnoses, diabetic medications or number of visits in the preceding year. All of these attributes have been chosen by clinical experts to be potentially associated with patient's diabetic condition.

As shown on Fig. 1a, 3 classes describing whether and when was patient readmitted are provided for prediction. According to the authors of the dataset [1] and our task definition, we transformed the problem to binary classification of *early readmission*. In this scenario, records of patients readmitted in less than 30 days are considered to be in positive class (early readmitted), otherwise not early readmitted. Final distribution of the classes after task adjustment is shown on Fig. 1b. As we can see from the figure, classes are highly imbalanced. We provide also detailed analysis of the dataset and its attributes².

3 Method proposal

Our method to help to solve the task of early readmission prediction is based on basic machine learning workflow, using variety aspects of data analysis and machine learning.

From data analysis, we have found out that most of the attributes are categorical. Depending on concrete attribute type (nominal or ordinal) correct encoding has to be chosen.

There are also attributes with a lot of missing values (4 attributes only), where we will have to choose whether fill the missing values and use the attribute for prediction or not. In other attributes, corrupted values have to be fixed (e.g. unknown gender). In numerical attributes, we do not find any extreme values or outliers (all values looks natural).

² <https://github.com/pmacinec/diabetes-patients-readmissions-prediction>

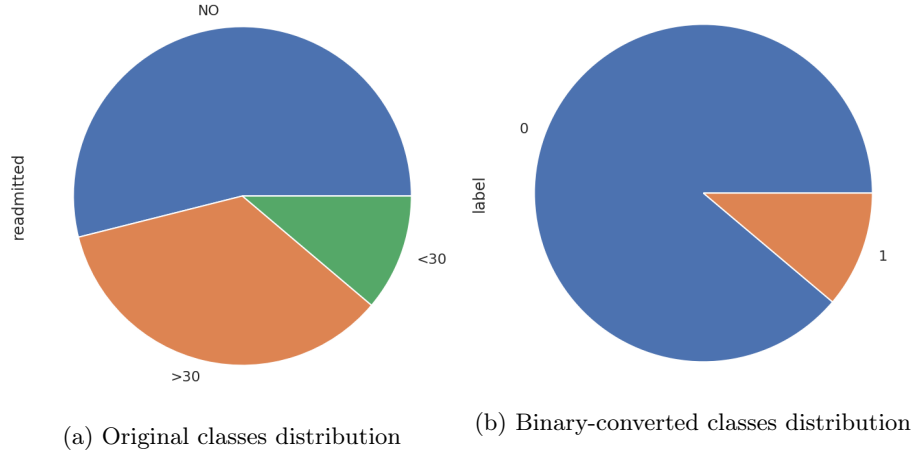


Fig. 1: Distribution of predicted classes - original (whether and when was patient readmitted) and after converting to binary form (only whether was patient readmitted or not).

Next challenging problem will be to handle imbalanced classes. Minor class has just about 10% of the values. According to chosen algorithm and preprocessing, we will use either undersampling or oversampling.

References

1. Strack, B., Deshazo, J., Gennings, C., Olmo Ortiz, J.L., Ventura, S., Cios, K., Clore, J.: Impact of hba1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed research international* **2014**, 781670 (04 2014). <https://doi.org/10.1155/2014/781670>