

Generalization of Deep Neural Network of Hospital Readmission Prediction Models for Diabetes Patients Using Apache Spark Clustering

Fatma Al-Rubaei, Mohammed Alhanjouri, Jashwanth Kalyan Polavarapu

October 23, 2022

1 The main idea

By relieving the strain on the front desk, concentrating resources on patients who require specialized care that could save their lives, being cost-effective, and ultimately increasing the patient's quality of life, reducing it would benefit healthcare services tremendously. readmission can be reduced by considering the factors of the patients age, race, number of diagnoses, and others. This relation can be studied by using different machine learning techniques such as unsupervised learning, which includes clustering algorithms such as k-means and supervised learning which includes regression and classification algorithms such as KNN, decision trees, and neural networks.

2 The methodology

The dataset used in this study was extracted from the Health Facts database, which is a national data warehouse that collects clinical records from hospitals across the United States, by Beata Strack in his study and donated to the UCI machine learning repository. The data set was compiled over a ten-year period (1999-2008) of clinical care at 130 US hospitals. It contains approximately 100,000 diabetes inpatient encounters with 55 different integer attributes.

3 The results

They tested a variety of predictive models, and it successfully and accurately predicted unplanned readmissions in diabetic patients. Our best and most promising model (DNN) achieved a training set accuracy of 66% and a testing set accuracy of 50%.

4 Recommendation

In Big Data and especially when we work with medical data, it is critical to reduce the execution time as much as possible. They report a 32% reduction in execution time when training machine learning algorithms with Apache Spark, indicating promising results for future work on big medical data analysis.