

# Predicting Hospital Readmission Patterns of Diabetic Patients using Ensemble Model and Cluster Analysis

Hung N. Pham, Anurag Chatterjee, Balasubramanian Narasimhan, Choon Wee Lee, Diksha Kumari Jha, Edric Yeng Fai Wong, Stella Ellyanti, Quang H. Nguyen, Binh P. Nguyen, Matthew C.H. Chua, Jashwanth Kalyan Polavarapu

October 23, 2022

## 1 The main idea

This paper proposes an ensemble model to predict hospital readmission by selecting from a pool of 15 models comprised of Logistic Regression, Decision Trees, Neural Network, and Augmented Nave Bayes networks.

## 2 The methodology

The dataset used in this study was obtained from the UCI machine learning repository, and it represents 10 years of clinical care at 130 US hospitals and integrated delivery networks from 1999 to 2008. Here, the methods used are

1. Performance Metrics
2. Predictive Modeling
3. Cluster Analysis using k-Means

## 3 The results

The ensemble classification accuracy was 63.5% on the 15% test partition of the data, which was higher than the baseline. When compared to the baseline SVM model, the ensemble has comparable accuracy but higher sensitivity. As higher sensitivity indicates that the model has a higher true positive rate, implying that it is better than the baseline model at predicting cases of re-admission.

## 4 Recommendation

The models were developed with a focus on the readmission of diabetic patients in the United States; however, the performance of these models in other geographies or for other chronic illness patterns, which have been shown to be dependent on different factors, has not been evaluated. In the future, we could train different prediction models, which could even be ensembles, to predict whether a patient is likely to be readmitted based on their membership in that cluster.