Thesis Defense

TWO-STEP PREDICTIVE MODEL FOR MISSED APPOINTMENTS AT OUTPATIENT PRIMARY CARE SETTINGS SERVING RURAL AREAS

Laith Abu Lekham

3:30 pm – 5:00 pm EST, Tuesday, June 9, 2020

Abstract

Missed appointments are a significant cause of inefficiency in the healthcare industry. Many researchers have studied this problem in various healthcare settings. However, few studies in the literature are concerned with predicting missed appointments at outpatient primary care settings serving rural areas. This study holistically investigates the factors behind two types of missed appointments – no shows and cancellations - at Finger Lakes Community Health (FLCH), an outpatient primary care medical center serving rural areas in New York, and develops a two-step predictive model that can be used to reduce their incidence. The dataset used in this research includes 40 features such as patient age, race, ethnicity, and lead time (the time between appointment creation date and actual appointment date). The label in this dataset represents the appointment status with two classes: show and missed appointments - no-shows and cancellations.

The study was carried out in three main phases. First, exploratory data analysis was conducted to discover the patterns related to the missed appointments. Also, a text mining framework was developed to conduct a root cause analysis (RCA) and Pareto analysis. Second, the association between some attributes and appointment status was analyzed using Chi-square and Welch's t-test. Third, a two-step predictive model for the missed appointments was built using machine learning algorithms: Multi-Layer Perceptron Artificial Neural Networks (MLP), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Decision Tree (DT), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (k-NN), and ensemble methods including Random Forest (RF), Ada-Boost (AB), Gradient Boosting (GB), and Bagging. The first step of the predictive model is designed to predict the missed appointments in more than one-week notice without using the weather forecasts (long term predictions). The second step of the model is designed to predict the missed appointments in less than a one-week notice by incorporating the weather forecasts in the model (short term predictions). Weather forecasts are only used in the first step model as they are usually

accurate in the short term. The multi-stage model is used to provide more accurate timely predictions and performance.

It was found that appointment lead time is a key driver for missed appointments. The longer the lead time, the more likely a patient is to miss an appointment. Also, the missed appointment rate decreases significantly as the day progresses. Based on the text mining framework for RCA and Pareto analysis, it was found that most of the missed appointments are related either to the patient or processes. The five most common causes that contribute to approximately 80% of the missed appointments are (1) personal issues or emergency, (2) forgetfulness, (3) sickness, (4) miscommunication, and (5) call center system issues. Association analysis showed that most of the attributes such as lead time, age, and care type are associated with the appointment status (*P*-value less than 0.05).

The ensemble methods and MLP performed the best among all the classifiers in predicting the missed appointments after tuning their hyperparameters with an average accuracy of more than 93% and 84%, respectively. Also, most of the classifiers showed moderate variability in their performance (via 10-fold cross-validation) in predicting missed appointments. Incorporating the weather as an external variable has improved the performance of the model significantly with an average best accuracy of 99%. Also, adding the weather variable improved the stability of the performance of most of the classifiers by reducing the variability (low standard deviation – less than 6%). Based on this analysis, some recommendations and interventions were proposed to reduce the missed appointments rate at FLCH including reducing appointment lead time, establishing a recall system, changing the walk-in hours, using the predictive model in scheduling patients, and improving communication and patient education.