Using Machine Learning to Identify No-Show Telemedicine Encounters in a New York City Hospital

Wanting Cui, Joseph Finkelstein

Center for Biomedical and Population Health Informatics

Icahn School of Medicine at Mount Sinai, New York, USA

ICIMTH 2022

Introduction

No-show visits

- Patients making an appointment with the healthcare centers, but failing to attend their appointments without previous notice.

A common and important problem for hospitals not only in the United States but several countries around the world

It could cost a major hospital over 15 million dollars annually

Methods to prevent no-show visit

- Reminder system
- Imposing penalization

The average no-show rate for a healthcare center was 3% to 18%

Introduction

Building predictive models to identify potential no-show patients

Current models [1]:

- Regression Models: Logistic regression, multiple linear regression
- Train Based Models: Decision trees
- Neural Network, Marko Based Models, Bayesian Models

All studies are in-person visits

Telemedicine visits are different:

- Less transportation constraint
- Higher requirements for technology

[1] Carreras-García D, Delgado-Gómez D, Llorente-Fernández F, Arribas-Gil A. Patient no-show prediction: A systematic literature review. Entropy. 2020 Jun;22(6):675

Objective

Build machine learning models to identify potential no-show patients in telemedicine visits

Identify significant factors that affect no-show visits

Method

Dataset

- Extracted from the electronic health record (EHR) at Mount Sinai Health
- Date: March 2020 to December 2020
- Telemedicine visits:
  - Video visits
  - Telehealth
  - Telephone visits
  - Telemedicine visits
  - Non-face to face visits

Method

The dataset was separated into two groups:

- Patients that didn’t show up for the visit
- Patients presented at the visit

We identified 10 factors that could be obtained prior to their arrivals

- Visit type
- Age, Sex, Race
- 5 New York City Boroughs
- Health providers’ primary specialty, providers’ type
- Day of the week
- Number of previous telemedicine visits and number of previous no-show encounters

Since each patient could have multiple encounters, we treated each encounter independently

Predictive Models

Dataset characteristics:

- There were over 257,000 telemedicine sessions
- Around 5,000 of telemedicine session were no-show encounters (2%)
- Imbalanced dataset

In our previous study, we explored the effectiveness of logistic regression and tree based models on imbalanced medical data prediction [1]

Tree based model with sampling achieved the best result

[1] Cui W, Bachi K, Hurd Y, Finkelstein J. Using Big Data to Predict Outcomes of Opioid Treatment Programs. Stud Health Technol Inform. 2020 Jun;272:366-369

Predictive Models

Machine learning models:
- Support vector machine (SVM)
- Random Forest (RF)
- Extreme gradient boosting (XGB)

Sampling on the training set:
- Radom up sampling
- Random under sampling
- Synthetic minority oversampling technique (SMOTE)

Parameter tuning, cross validation

Evaluation metrics: Area under the ROC curve (AUC)

Results

There were 257,293 telemedicine sessions between March 2020 and December 2020

5,124 of telemedicine session were no-show encounters (2%)

There were 152,164 unique patients in the dataset

4,150 patients had at least one no-show encounter during this time period (2.7%)

Results

10 predictors

Target variable (binary): whether a patient presented to the telemedicine session

Model	Sampling	CV AUC	Test Accuracy	Test AUC
SVM	Under	0.70	0.75	0.64
RF	Under	0.68	0.81	0.66
XGB	Under	0.68	0.74	0.68

Results

Investigated the feature importance of XGB model

Identified the top 5 factors:

- Patients' previous no-show encounters
- Race
- Boroughs
- Providers' type
- Providers' specialty

Table 2. Top features affecting patients' no-show rate based on patients' information

	No Show Encounters			Present Encounters	
	count	percent		count	percent
Previous no show					
0 times	4171	81.40%		245999	97.60%
1-2 times	605	11.80%		5142	2.00%
3 or more times	348	6.80%		1028	0.40%
Race					
Asian	269	5.20%		15126	6.00%
Black	1077	21.00%		31392	12.40%
Others	2253	44.00%		87517	34.70%
White	1525	29.80%		118134	46.80%
Borough					
Bronx	658	12.80%		18916	7.50%
Brooklyn	757	14.80%		42537	16.90%
Manhattan	2155	42.10%		87279	34.60%
Others	923	18.00%		75508	29.90%
Queens	631	12.30%		27929	11.10%

Table 3. Top features affecting patients' no-show rate based on providers' information

	No Show Encounters			Present Encounters	
	count	percent		count	percent
Provider Type					
Nutritionist	163	3.20%		1817	0.70%
Physician	3382	66.00%		206488	81.90%
Psychologist	157	3.10%		4171	1.70%
Social Worker	707	13.80%		8600	3.40%
Provider Specialty					
Cardiology	81	1.60%		10479	4.20%
Dermatology	106	2.10%		10912	4.30%
Endocrinology	137	2.70%		16455	6.50%
Nutrition	163	3.20%		1223	0.50%
Pediatric care	141	2.80%		13091	5.20%
Adult Psychiatry	472	9.20%		6475	2.60%
Children Psychiatry	319	6.20%		2383	0.90%

Discussion

XGB was the best model, it had the highest AUC score

XGB model could provide feature importance that allowed us to analyze factors that are associated with no-show encounters

Patients with previous no-show encounters, non-White or non-Asian patients were important factors for no-show visits

Patients’ location (Borough) was an import factor

- Patients do not need to travel to hospital or clinics
- Related to patients’ socioeconomic factors

In future studies:

- Explore more machine learning and sampling methods to increase the prediction accuracy
- Map Zip code into income level, education level and other socioeconomic factors

Conclusion

XGB with under sampling was the best machine learning model to identify no-show patients using telemedicine service

Patients' previous no-show encounters, race and location (boroughs), providers' type and specialty were the 5 factors that were highly correlated to no-show encounters

Physicians with specialities in psychiatry and nutrition, and social workers were more susceptible to higher patient no-show rate

Thank You!

CONTACT INFORMATION:

WANTING CUI

EMAIL: WANTING.CUI@MSSM.EDU