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

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

Predictive Models

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

Sampling on the training set:
- Random 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 sessions 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Sampling</th>
<th>CV AUC</th>
<th>Test Accuracy</th>
<th>Test AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>Under</td>
<td>0.70</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>RF</td>
<td>Under</td>
<td>0.68</td>
<td>0.81</td>
<td>0.66</td>
</tr>
<tr>
<td>XGB</td>
<td>Under</td>
<td>0.68</td>
<td>0.74</td>
<td>0.68</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>No Show Encounters</th>
<th>Present Encounters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>count</td>
<td>percent</td>
</tr>
<tr>
<td><strong>Previous no show</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 times</td>
<td>4171</td>
<td>81.40%</td>
</tr>
<tr>
<td>1-2 times</td>
<td>605</td>
<td>11.80%</td>
</tr>
<tr>
<td>3 or more times</td>
<td>348</td>
<td>6.80%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>269</td>
<td>5.20%</td>
</tr>
<tr>
<td>Black</td>
<td>1077</td>
<td>21.00%</td>
</tr>
<tr>
<td>Others</td>
<td>2253</td>
<td>44.00%</td>
</tr>
<tr>
<td>White</td>
<td>1525</td>
<td>29.80%</td>
</tr>
<tr>
<td><strong>Borough</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronx</td>
<td>658</td>
<td>12.80%</td>
</tr>
<tr>
<td>Brooklyn</td>
<td>757</td>
<td>14.80%</td>
</tr>
<tr>
<td>Manhattan</td>
<td>2155</td>
<td>42.10%</td>
</tr>
<tr>
<td>Others</td>
<td>923</td>
<td>18.00%</td>
</tr>
<tr>
<td>Queens</td>
<td>631</td>
<td>12.30%</td>
</tr>
</tbody>
</table>
Table 3. Top features affecting patients’ no-show rate based on providers’ information

<table>
<thead>
<tr>
<th>Provider Type</th>
<th>No Show Encounters</th>
<th>Present Encounters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>count</td>
<td>percent</td>
</tr>
<tr>
<td>Nutritionist</td>
<td>163</td>
<td>3.20%</td>
</tr>
<tr>
<td>Physician</td>
<td>3382</td>
<td>66.00%</td>
</tr>
<tr>
<td>Psychologist</td>
<td>157</td>
<td>3.10%</td>
</tr>
<tr>
<td>Social Worker</td>
<td>707</td>
<td>13.80%</td>
</tr>
<tr>
<td>Provider Specialty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiology</td>
<td>81</td>
<td>1.60%</td>
</tr>
<tr>
<td>Dermatology</td>
<td>106</td>
<td>2.10%</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>137</td>
<td>2.70%</td>
</tr>
<tr>
<td>Nutrition</td>
<td>163</td>
<td>3.20%</td>
</tr>
<tr>
<td>Pediatric care</td>
<td>141</td>
<td>2.80%</td>
</tr>
<tr>
<td>Adult Psychiatry</td>
<td>472</td>
<td>9.20%</td>
</tr>
<tr>
<td>Children Psychiatry</td>
<td>319</td>
<td>6.20%</td>
</tr>
</tbody>
</table>
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