Medical Appointments No Shows

Abe Durrant

**Introduction**

The goals of this analysis were to predict a patient’s probability of a no-show at an appointment as well as to make recommendations to the operation unit based on discoveries in the data. A dataset including patients’ medical history as well as personal characteristics was used in the analysis.

In addition, I used the factors in the dataset to create new features of the data to use in the analysis. I looked at how many days there were between a scheduled appointment and the appointment day, what day/month/hour the appointments were scheduled at and held. I also looked at if the patient had been a no-show in the past.

**Analysis**

Before fitting the predictive model, a few relationships were looked at. Age seems to be connected with a patient’s probability of a no-show as middle-aged patients tend to miss less appointments while patients 18-25 and above 65 tend to miss more appointments.

Chart, line chart

Description automatically generated

The time that goes by after scheduling the appointment and before the appointment happens also seems to be connected with being a no-show. Appointments after 15 days seem to be higher in the number of no-shows than appointments before 15 days have passed.

Chart, line chart

Description automatically generated

**Model**

In order to predict a patient’s probability of a no-show a Random Forest model was used. This is a machine learning model that uses decision trees to predict the probability of a patient missing. It outperformed other models that were tested on the data and was approximately 80% accurate when predicting new data. It can return probabilities that a patient will no-show. It was able to predict about 63% of no-shows when an appropriate probability threshold was set.

**Results**

The following are the most important factors in the model in determining a no-show:

Chart, bar chart, histogram

Description automatically generated

We can see that age is important in determining whether a patient will show up. As seen before, patients 18-25 and older patients are the patients more likely to miss. We also see that the time since scheduling is a big factor. The day of the week is also a factor and Friday and Saturday see a higher percentage of missed appointments than the rest of the week.

From these discoveries I recommend that the units contact patients with a text reminder when the model gives them a higher probability of a no-show. Text messages tend to work best with patients who have scheduled under 90 days out. Patients who are between 18-25 or patients with Friday and Saturday appointments should also be contacted more often with a reminder.

Additionally, since text messages may work better when patients have scheduled within 90 days, I recommend that a new form of contacting patients before an appointment be tested on patients that have scheduled further out (such as a phone call or email) to see if they better respond. It is primarily older patients that schedule this far out so a different form of communication might be more effective. There is little data on patients scheduling this far in advance, so it is possible that texts will benefit them as well.

Overall, the model can be used to predict which patients have higher probabilities of no-showing. When a patient has a higher probability of a no-show a text message reminder should be used.