

**Lowering the No-Show Rates of**

**Clinical Appointments**

**Through Patient Identification and**

**Targeted Text Reminder Intervention**

**Team 3**

Eleanor Lee

Yifei Ren

Howard Wang

Julie Wang

**TABLE OF CONTENTS**

[1 Business Understanding 3](#_Toc37625509)

[1.1 Background 3](#_Toc37625510)

[1.2 Previous Studies 3](#_Toc37625511)

[1.3 Our Purpose 4](#_Toc37625512)

[2 Data Understanding 6](#_Toc37625513)

[3 Data Preparation 10](#_Toc37625514)

[3.1 Feature Engineering 12](#_Toc37625515)

[3.2 Encoding 12](#_Toc37625516)

[4 Modeling: Classification Model 13](#_Toc37625517)

[4.1 Rebalancing Training Data 13](#_Toc37625518)

[4.2 Model Comparison 14](#_Toc37625519)

[4.3 Model Optimization 14](#_Toc37625520)

[4.4 Model Results 14](#_Toc37625521)

[4.5 Feature Importance 15](#_Toc37625522)

[4.6 Preparing Dataset for Recommender System 17](#_Toc37625523)

[5 Modeling: Recommender System 17](#_Toc37625524)

[5.1 Preparing Dataset for Recommender System 18](#_Toc37625525)

[5.2 Similarity Measure and Cutoff Value 18](#_Toc37625526)

[5.3 Patients Recommended to Send Text Reminder 19](#_Toc37625527)

[6 Evaluation 19](#_Toc37625528)

[6.1 Classification 20](#_Toc37625529)

[6.2 Recommender System 21](#_Toc37625530)

[7 Deployment 24](#_Toc37625531)

[8 Team Member Contributions 25](#_Toc37625532)

[9 Bibliography 26](#_Toc37625533)

[10 Appendix/Screenshots 27](#_Toc37625534)

[10.1 Understanding Appointment Distribution 27](#_Toc37625535)

[10.2 Histograms of Variable Distribution & Correlation 28](#_Toc37625536)

[10.3 Classification Models 31](#_Toc37625538)

[10.4 Recommender System 36](#_Toc37625539)

# 1 Business Understanding

## 1.1 Background

The healthcare industry in the United States is growing incredibly fast. Under current administration rules, the national healthcare spending is estimated to grow about 5.5% per year from 2018-2027. At that rate, the industry is expected to be worth about $6.0 trillion by 2027. Healthcare expenditures related to physician and clinical services was worth about 725.6 billion in 2018, growing 4.1%. Each person will spend about $11,172, which is 17.7% of gross domestic product[[1]](#footnote-1).

Given how the industry bills time, procedures, and prescriptions, a doctor’s time can be analogous to revenue. If he has a patient scheduled, and that patient is a no-show to that appointment, that is a ‘wasted’ time slot where potential revenue was lost. In 2011, the average cost of a doctor’s appointment (length of 15 minutes) was $104, with $69 paid by the patient and the rest by insurance in the United States[[2]](#footnote-2). A doctor at a clinic who works a typical 8 hour day could see, on average, 32 patients a day, totaling about $3,328 in revenue. Not only is revenue lost, but those patents that are walk-ins may have to wait even longer (which could be detrimental to their health if it is urgent) or may leave and go to another clinic (resulting in more loss revenue)

## 1.2 Previous Studies

This kind of study has been done before by biohealth informatics researchers at various universities like Duke and Johns Hopkins.

Johns Hopkins developed a model, used by 2 clinics since 2017, that assigned ‘no-show scores’ to patients so that clinics could identify patients that had a high-risk of not showing up. They also discovered that patients who frequented emergency departments more often were also more likely to miss their scheduled appointments. In addition, patients who used online patient portals for scheduling, were less likely to miss their appointments[[3]](#footnote-3).The implementation of their model allowed those two clinics to focus on outreach strategies or give those slots to someone who needed care immediately. Researchers at Duke experimented with their models by seeing how they could increase the accuracy by either widening the scope of the or limiting the model specifications for each clinic[[4]](#footnote-4). They found that at the clinic-level models had both a higher accuracy, and higher recall.

While Duke’s study has shown that clinic-level models can be more accurate, due to a limitation of data, we will be doing a general model across clinics in order to learn more from our dataset. Our dataset is in Brazil, which provides universal health care, but the insights of this model are applicable to U.S. clinics as well. By training a model to identify at-risk patients, we can learn what features are informative and help cut costs to clinics that are already running at almost full capacity that come from wasted schedule slots.

## 1.3 Our Purpose

Our goal is to identify patients that are likely to be no-shows. We want to identify these patients for 2 reasons: appointment reminders and slot efficiency. There are patients who are less likely to be no-shows if they are simply reminded by a text message. In addition, if a patient is likely a no-show to their appointment, the clinic could usher in a walk-in patient to take up that appointment slot instead. By identifying those patients that are likely no-shows, clinics can reach out to remind patients about appointments as well as maximize their slot efficiency for walk-in patients.

Our model will also take it a step further. Aside from identifying patients who will be no-shows, we will create a user-based recommender system to help clinics maximize their outreach efforts. We do not want to waste money and time texting patients that are still likely to be ‘no-shows’, nor do we want to text people that are already predicted to show up. We will be using our predictions from our first model to identify patients that are predicted ‘no-shows’. Then, we will see which patients have not been texted, and find similar patients that showed up and were texted. We will be able to best direct a clinic’s efforts to patients that are likely to become ‘shows’ after being texted.

We believe that this recommender system would be an effective next step to attempt in order to decrease clinic no-shows, as there is previous work showing the effectiveness of text reminders. A study published in the International Journal of Pediatrics focused on increasing appointment adherence in a pediatric clinic serving primarily low-income, African American families. The preliminary survey data for the study found that most preferred texts 3 days before the scheduled appointment. However, the texts could not be created using automated texting programs due to HIPAA concerns. The researchers manually reminded patients in the intervention group 3 days before their scheduled appointment. The study concluded that the no-show rate was 23.5% in the intervention group, compared to a 38.1% no-show rate in the control group; people in the intervention group were 2 times more likely to show up to their appointments, or at least call to cancel/reschedule as necessary[[5]](#footnote-5).

This study shows that sending text reminders is an effective way to lower the no-show rate at a clinic (applicable to other demographics, as the study also found that no demographic factors affected the effectiveness of text reminders on the no-show rate). However, an issue this study revealed is that automated text messages are not HIPAA compliant, so manual text messages must be sent and drafted from dedicated phones, and issue when labor is limited. Our recommender system would recommend the patients that are more likely to respond to the intervention method, so that the clinic can focus their efforts efficiently. That is why we will take the next step further to identify affectable patients instead of suggesting the clinic mass-text those that are predicted to be no-shows.

# 2 Data Understanding

Our data comes from Kaggle in a csv form. It contains appointment data from various Brazilian clinics. There are a total of 110,527 observations(appointments) of 14 variables (13 features and the target variable) scheduled between November 2015 and June 2016, and appointment dates between April 2016 and June 2016. Here is a table describing each variable, its type, and any additional context that may be necessary:

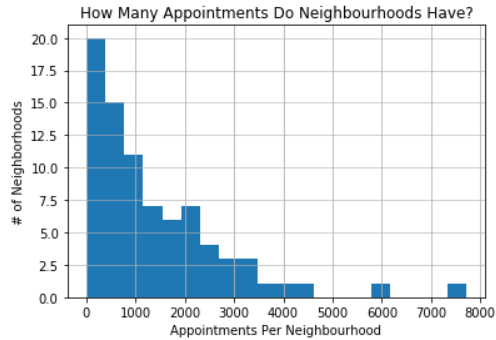
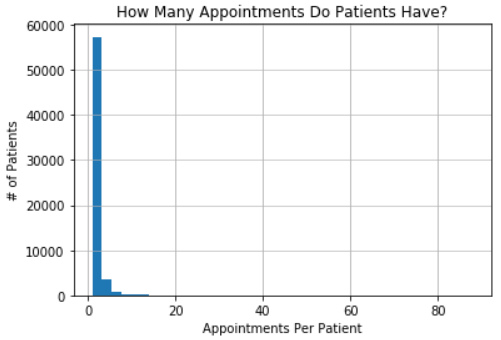
|  |  |  |
| --- | --- | --- |
| Variable Name | Type | Description/Purpose |
| PatientId | Integer | Identification of a Patient |
| AppointmentID | Integer | Identification of each Appointment |
| Gender | Categorical | Gender of the Patient |
| ScheduledDay | Datetime | Day/Time someone called or registered the appointment |
| AppointmentDay | Date | Day of the Actual Appointment |
| Age | Integer | How old is the patient |
| Neighbourhood | Categorical | Where the Appointment takes Place (e.g. where the clinic is located) |
| Scholarship | Binary\* | Whether the patient is a part of the Brazilian social welfare program,Bolsa Família, where they receive financial aid and children are required to receive vaccines and go to school (<https://en.wikipedia.org/wiki/Bolsa_Fam%C3%ADlia>) |
| Hipertension | Binary\* | Whether the patient has abnormally high blood pressure |
| Diabetes | Binary\* | Whether the patient has diabetes |
| Alcoholism | Binary\* | Whether the patient has alcoholism |
| Handcap | Binary\* | Whether the patient has a handicap |
| SMS\_recieved | Binary\* | 1 or more messages was sent to this patient |
| No-show | Categorical\*\* | Is the patient a no-show |

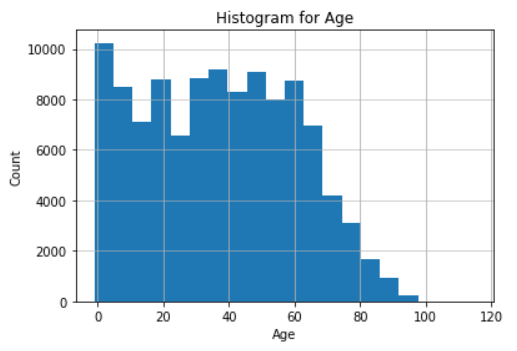
\*All binary variables have 1 indicating TRUE

\*\* The target variables has values of ‘Yes’ or ‘No’

As we examined our data, we wanted to understand how the observations were spread out amongst the demographic and health variables. We also used this as an opportunity to examine anomalies that may need to be addressed in the data cleaning process.

There were no N/A or Null values in any of the columns. While there were 110,527 appointments, these were all made by 62,299 patients. Examining the cumulative distribution, we noticed that about 80% of appointments were done in 40% of the neighborhoods, and 80% of the appointments were held by 65% of patients.



There are 81 neighbourhoods in this dataset. Unfortunately, there is no information about how many clinics are operating in each neighbourhood. Most neighbourhoods had less than 1000 appointments, however there are a few neighborhoods serving over 4000 appointments. The busiest neighborhood is Jardim Camburi with 7717 appointments. The least busy neighborhood is Parque Industrial with only 1 appointment. If we look further at the breakdown of patient-related variables, we notice some interesting facts and issues.

For Age, we found that there was an observation of -1, an erroneous input. The maximum age was 115, while the minimum was 35. Looking at the distribution of ages, we see a slight dip for the teens, 20s, and 30s, plus a gradual tail at the upper range as people slowly die off. This makes sense as those in the teens, 20s, and 30s are less likely to need to come in for routine checkups or other follow ups as they will have less health issues. The rise at 20 could be due to health issues that tend to surface after puberty, but before middle age, like lupus, multiple sclerosis, and various skin disorders.

For Scholarship, most patients were not on scholarship. Only about 9.83% of appointments had patients who were on scholarship, receiving social welfare and requiring children show up to receive vaccinations. Of appointments with patients on scholarship, 23.73% were no-shows, compared to 19.81% no-show rate of those not on scholarship.

For Hipertension (hypertension, relating to high blood pressure), the majority of appointments were patients not suffering from hypertension. 19.72% of the appointments are for patients with hypertension. It seems that appointments belonging to those suffering from high blood pressure are less likely to miss appointments; 17.35% no-shows compared to 20.77% of those that do not have hypertension.

Another medical variable is diabetes. Only 7.19% of the appointments have patients with diabetes. Those appointments with patients that have diabetes are slightly less likely to be no-shows; 18% compared to 20.36%.

Another variable is Alcoholism. There is a low proportion of appointments scheduled by patients suffering from alcoholism: 3.04%. There is a negligible difference in no-show rates of appointments that have alcoholic patients and non-alcoholic patients.

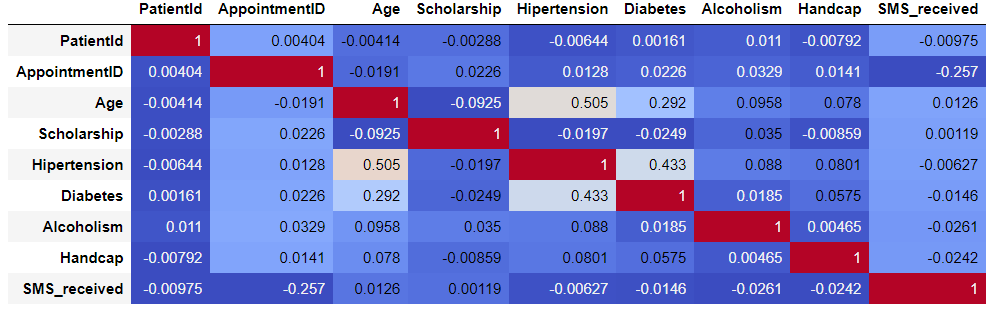
Handicapping is interesting. While the data dictionary indicates that Handicap is a binary variable, there are values of 2, 3, and 4. These observations take up less than 1% of the observations, and are therefore most likely errors, which we will ignore in this analysis. About 1.85% of appointments deal with patients that have a handicap. For appointments dealing with patients that have handicap, there is a no-show rate of 17.92%, compared to 20.23% for appointments of patients with no handicap.

SMS texts were sent to a little over a third of all appointment patients. Interesting enough, the rate of no-show for appointment patients that received a reminder text is 27.6%, much higher than those who did not (16.7%). This is possibly due to ineffective outreach, skewed by clinics who experience high no-show rates and are trying to combat that by sending reminders to every single appointment.

The majority of appointments have female patients (65%). However, there seems to be no difference in the rate of no-show between male and female appointments.

Overall, the no show rate is about 20.19% across all appointments. From a preliminary examination, it seems that useful patient-level variables for our model may be Scholarship, Diabetes, Hipertension, Handcap, and SMS\_received.

Finally, we examined the correlation of the variables to see if there is multicollinearity.

****

We found no variables that were extremely correlated. Thus, we did not remove any variables as none seemed redundant.

# 3 Data Preparation

In order to create a functional and realistic model, we had to remove some clearly erroneous inputs. We deleted an observation that had the age of the child as -1. We also deleted observations where the Handicap was 2,3, or 4, even though the variable is defined as a binary 0,1 variable.

We also had to make several decisions in how to interpret whether the data was wrong or not. There were several observations where prepubescent children were marked as alcoholics. However, Brazilian regulation and culture surrounding alcohol is different from the US. According to SciElo, a Brazilian Psychiatry journal, a paper on adolescent alcoholics stated the mean age to be 14[[6]](#footnote-6). So, we decided to only remove acholic observations where the age was below 10.

Another problem with observations dealing with children was deciding whether to eliminate observations where children were too young (less than 14 years old) to take themselves to the clinic. This would be a problem for the recommender system portion if these children were not eligible to be sent SMS’s. However, after looking at the data further, there are younger children who ‘received’ SMS’s, so we will assume these messages were sent to a guardian, presumably the one that would be responsible for setting the appointment and accompanying them to the clinic.

We presumed that time would, logically, be an important feature in this model, due to the nature of the problem. The longer between scheduling and appointment, the more people would miss; people would be more likely to skip weekday appointments due to work; etc. The granularity of the time component needed to be determined, whether we would look at it in seconds, minutes, hours, days, etc. The data come in the form of Y-M-D H:M:S, for example, ‘2016-04-29T17:29:31Z’. However, we don’t need all this information, especially in this form. So, we edited the date format to be ‘Y-M-D’, for example, ‘2016-04-29’

We removed any observations where the Scheduled Day happened after the Appointment Day. We also decided to remove the timestamp so that we can focus on ‘Day’ as our level of granularity. Instead of having to resort to hours or minutes for observations that may only have hours between the scheduling and the appointment time, we can just have the difference be 0 days, so that is more understandable. This is especially important for appointments that may have been scheduled over a month in advance. We also deleted any observations where the AppointmentDay is recorded to be before the ScheduledDay, which is impossible.

We adjusted our target variables to be binary; switching from values of Yes/No to 1/0 for the purpose of modeling clarity. We are left with a cleaned dataset of 71,834 observations/unique appointments after removing observations based on the discussions above.

## 3.1 Feature Engineering

After anomalies were removed and formatting adjusted, we moved onto engineering new features that may prove to be significant, as well as encoding others so that we could input them into our model.

We created several variables, described here in this table:

|  |  |
| --- | --- |
| Variable Name | Description & Calculation |
| Lead\_Time (Integer) | Difference between ScheduledDay and AppointmentDay |
| Scheduled\_Weekday (Binary) | Whether the scheduling was done a weekday or not |
| Appt\_Mon (Binary) | Whether the appointment is on Monday |
| Appt\_Tues (Binary) | Whether the appointment is on Tuesday |
| Appt\_Weds (Binary) | Whether the appointment is on Wednesday |
| Appt\_Thurs (Binary) | Whether the appointment is on Thursday |
| Appt\_Fri (Binary) | Whether the appointment is on Friday |
| Appt\_Sat (Binary) | Whether the appointment is on a Saturday |
| Appt\_Sun (Binary) | Whether the appointment is on a Sunday |
| Neigh\_No-show\_Percentage (Float64) | The No-Show rate at in that neighborhood |

## 3.2 Encoding

The only categorical variables in this dataset are Gender and Neighbourhood. We used One Hot Encoding for this data. Gender was easy to encode, as the dimensionality remained the same. Gender was encoded to be ‘is\_male’, where 1 indicates a male patient.

Encoding the clinics was a challenge. With 81 neighborhoods, the dimensionality of the data would expand to have an additional 80 columns. However, since we also have the Neigh\_No-Show\_Percentage, we decided that it may not be necessary to have all 80 neighborhoods encoded. From our data exploration earlier, we found that 40% of clinics, about 33 clinics, accounted for 80% of the appointments. With this in mind, we decided to encode the 33 busiest neighborhoods in the form of ‘is\_NeighbourhoodName’, with a 34th column called ‘Other’ for the remaining 48 neighborhoods. We believe this variation of encoding will help our models learn the most without extending the width of our dataset too much.

Both original columns ‘Gender’ and ‘Neighbourhood’ were dropped from the dataset after encoding was done.

# 4 Modeling: Classification Model

The classification models we will be testing are Decision Tree, Logistic Regression, kNN, SVM, and CatBoost. Due to the nature of some of these models, it was necessary to balance the training dataset and normalize the feature values. We split our data into Train and Test datasets, with a 70/30 split: 50, 283 in Train and 21551 in Test. Our modeling was done in using various packages from Python, like scikit-learn and catboost.

## 4.1 Rebalancing Training Data

Our dataset’s target variable, No-Show, is unbalanced, with 20% of observations being positive for No-Show. We plotted a learning curve based on a default Logistic Regression classifier. We found that we have sufficient training data, but we still decided to oversample to balance the dataset because we did not want to lose any potential information from the minority positive observations. Our Train dataset grew to 71,888 observations, with each outcome having 35,944 observations.

## 4.2 Model Comparison

In order to know how to approach this problem and which algorithm may turn out to be the best, we have outlined the pros and cons of each algorithm in the table below.

|  |  |  |
| --- | --- | --- |
| Model | Pros | Cons |
| Decision Tree | - Easy to implement  - Intuitive to explain | **-** Creates deep, unbalanced trees due to one-hot encoding |
| Logistic Regression | - Provides probabilities, so that a cutoff can be implemented  -Works well with diagonal decision boundaries | -High bias |
| kNN | -No assumptions about data  -Easy to understand and explain | -Lazy algorithm; time consuming  -Computationally expensive |
| SVM | -Performs well with linear and non-linear boundaries depending on the kernel used  -Handles high dimensional data well | -Susceptible to overfitting issues based on kernel |
| CatBoost | -Learns sequentially and builds off previous trees  -Can handle categorical features well | -Can overfit if trees are too large |

## 4.3 Model Optimization

The method we used for all the models was a grid search using cross folds validation with a k of 5. This allowed us to tune certain parameters as well as find the best estimator for each model. The cross folds validation aspect will help ensure that we use all our available data to inform our model making.

## 4.4 Model Results

We are focused on how well we can identify patients of the positive class, indicating that

the patient was a no-show to the appointment. We looked at metrics like precision, recall, accuracy, and AUC.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Precision | Recall | Accuracy | AUC | Additional Notes |
| Decision Tree | .34 | .36 | .62 | .54 |  |
| Logistic Regression | .34 | .57 | .57 | .59 |  |
| kNN | .36 | .3 | .65 | .57 |  |
| SVM | .34 | .59 | .56 | .57 |  |
| CatBoost | .71 | .90 | .86 | .87 | Used non-normalized and unbalanced training set |

We see that the Decision Tree, Logistic Regression, and SVM performed similarly on precision. However, Logistic Regression performed much better with the recall. Since SVM used a linear kernel, it’s not surprising to see that it has almost the exact same statistics as the Logistic Regression model.

Clearly, the best model is the CatBoost model. It not only has the best performance for any of the 4 metrics, but it also has a recall rate of .9. The model would be able to identify about 90% of no-show patients with a 71% precision, allowing clinics to identify most of their patients that will be no-shows without misidentifying too many patients that would be shows. This high precision will allow clinics to focus their resources and outreach methods wisely and to greater, measurable effect.

## 4.5 Feature Importance In order to understand why the CatBoost model performed so well, we looked at the feature importance. We pulled out the weights for each feature, and below are the features that had any weight.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Feature | Weight |  | Feature | Weight |
| Age | 30.001647 |  | is\_ITARARÉ | 1.291575 |
| Lead\_Time | 24.334057 |  | is\_ILHA\_DO\_PRÍNCIPE | 0.793934 |
| Neigh\_No-Show\_Percentage | 15.588332 |  | is\_NOVA\_PALESTINA | 0.512376 |
| SMS\_received | 11.548552 |  | Appt\_Tues | 0.425326 |
| Scholarship | 5.9321338 |  | is\_DA\_PENHA | 0.404885 |
| is\_Male | 3.3797398 |  | Scheduled\_Weekday | 0.216260 |
| Appt\_Weds | 3.3304383 |  | Appt\_Sat | 0.010610 |
| Appt\_Mon | 2.2301275 |  |  |  |

While we expected some features to be informative given the nature of the problem (Lead\_Time, Neigh\_No-Show\_Percentage, SMS\_received), it’s a surprise that Age is the most important variable.

Some of our earlier data understanding expectations are proven to be correct here, as Scholarship and is\_Male are the 5th and 6th most important features.

It is also surprising to see Appt\_Weds, Appt\_Mon, Appt\_Tues, and Appt\_Sat all have some weight. This indicates that people tend to miss appointments that are scheduled at the beginning of the week. This may be due to work, school, or simply forgetting as the new week starts. For those that miss Saturday, it may also be due to childcare, work, or waking up late on the weekend.

Another date-related feature of importance is Scheduled\_Weekday. It seems that scheduling on the weekend or weekday may affect if patients remember that they scheduled an appointment in the first place, maybe getting distracted before they can put the appointment in their calendar.

About 4 neighborhoods showed up in the list of important features. These 4 neighbourhoods are among the neighborhoods with the highest no-show rates, between 28%-32%.

## 4.6 Preparing Dataset for Recommender System

In order to prepare our predictions to be placed into the recommender system, we assigned each appointment to a group. The groups were predicted No-Show and No Text, predicted Show and Text, and Other. We are focused primarily on those that were predicted to not show and had not been texted to find how similar they are to those who showed and were texted, to see how likely they are to show if they are sent a text message. This will help the clinic maximize their outreach efforts and minimize costs.

The recommender system takes into consideration all the medical descriptive and demographic features to help calculate similarity.

# 5 Modeling: Recommender System

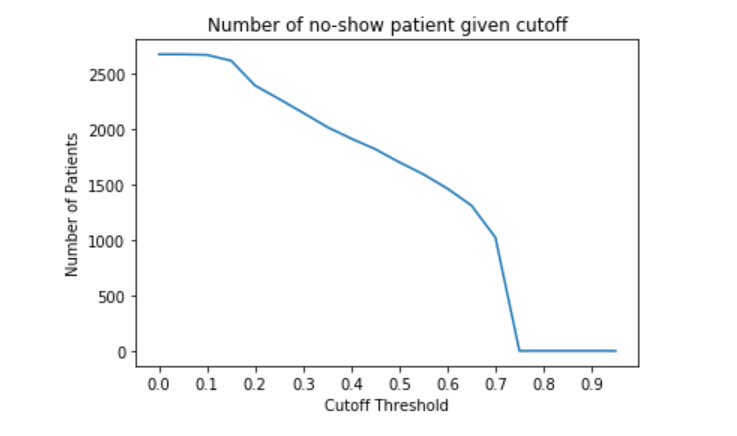
The recommender system we are building is a user-based collaborative filtering recommender system with implicit rating. Unlike the cases of purchasing products or rating a movie, the ‘rating’ we use here to construct similarity measures is the attribute a person possesses. For example, whether a person has hypertension, diabetes, or alcoholism are all attributes that a person has, and we are constructing cosine similarity on top of these features. Since our goal is to decide whom to recommend (send out a text reminder), then it is important to find the group of people who we predict to be no show patients, and for each patient in such group, find out the average cosine similarity to the group of people who received a text reminder and showed up. Our underlying assumption is that those who we predicted to be no-show and are very similar to the group of people who showed up after receiving a text reminder, will have a higher likelihood of showing up if we send them a text reminder as well.

## 5.1 Preparing Dataset for Recommender System

In our recommender system, we are using cosine similarity. Simply put, cosine similarity gives you the number of common attributes divided by the total number of attributes we have. Since cosine similarity is also subject to the variance across different features, we will have to normalize some of our continuous variables such as age and lead time, to (0,1) scale, and then we can use it along with our encoded dummy variables to calculate cosine similarity.

## 5.2 Similarity Measure and Cutoff Value

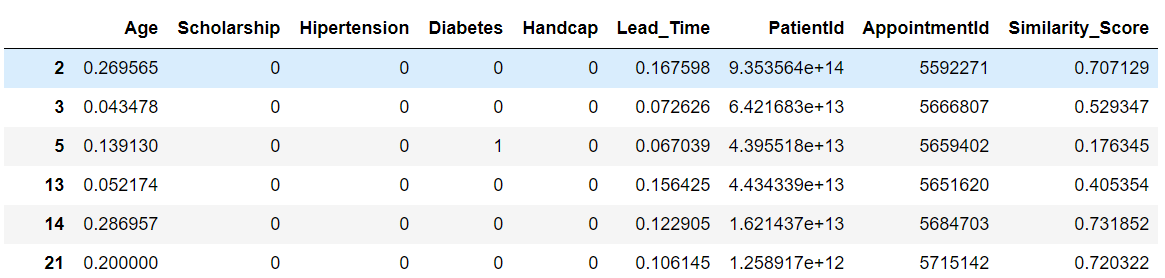
Based on the returned similarity and cutoff value we choose, we can also get a sense of the distribution of people qualifying for a text reminder given different cutoff value, which is shown below.



As we can see here, most of the people have a similarity score of 0.7 or less. We can use this distribution along with cost information of sending out a text to conduct cost-benefit analysis, which will be shown below.

## 5.3 Patients Recommended to Send Text Reminder

With the retrieved similarity score and a carefully chosen cutoff value, we can finally get the list of appointments we’d like to target by simply filtering the returned data frame on the similarity score column, and then send text reminders to the patients related to such appointments.

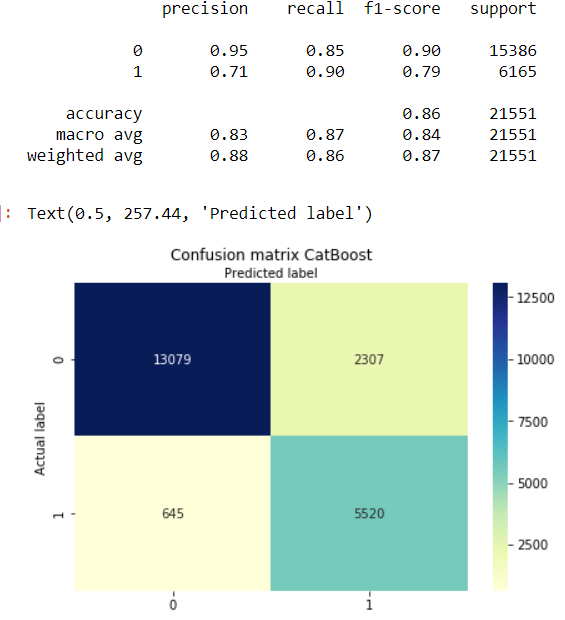


# 6 Evaluation

For the evaluation of the best classification model and recommender system, we will be using a few key assumptions. A systematic review found that across several studies of the effectiveness of text messaging reminders, the use of SMS reminders increased the likelihood of attendance at clinical appointments by **50%** We assumed that the SMS messaging service the clinics use is Bandwidth, a cheap alternative to Twilio. Bandwidth charges **$0.35** per phone line per month and charges **$0.005** per outbound text message. We also use the **$104** value mentioned in our research as a measure of ‘revenue’, as that is how much a missed doctor’s appointment costs the clinic, ignoring any walk-in patients that could be ushered in.

## 6.1 Classification

For the classification model, we will go ahead and assume that 50% of appointments that are reminded will convert to ‘Show’; we will consider how likely each patient is to convert in the recommender system evaluation. From our confusion matrix for the best classification model, the CatBoost model, we see that we have a total of 7827 patients predicted to be no-shows, with 2307 incorrectly predicted, and 645 not recalled. We assume that only half of those 5520 will become shows and texting people that would not have needed a reminder will not suddenly change them to be no-shows. Our data is also over the course of 2 months, so we will have a fixed cost of $0.70 for the phone line.



If the clinic has no budget restrictions, they can choose to remind all patients, which is the assumption we will make in this scenario as we cannot identify which patients a text message reminder would be or not be effective enough to flip them to show up to their appointment. This leads us to our calculations below:

|  |  |
| --- | --- |
| **Profit Recovered from Patients Predicted to be No-Shows Correctly** | 5520 \* .5 \* ($104 - $0.005) = **$287,026.20** |
| **Costs of Contacting Patients that were Show**  **(Predicted No-Show incorrectly)** | 2307 \* $0.005 = **$11.54** |
| **Costs of Contacting No-show Patients that still are No-Show** | 5520 \* .5 \* .005 = **$13.80** |

Based on these totals, by texting these patients just off the prediction of the classification model, we have a final equation of $287,026.20 - $11.54 - $13.80 - $0.70 = $287, 000.16. That is a recovered profit of $287,000.16 over these 2 months, as opposed to the clinic losing $574,000 (from $104 \* 5520) in revenue due to predicted no-shows not being given intervention. That is a recovery of about 50%.

## 6.2 Recommender System

Because we do not have any ‘test’ set to evaluate how our recommendation system performs, we will rely on the metric we mentioned earlier to provide a preliminary, hypothetical evaluation. The classification model has an assumption of 2x likelihood, or 50% conversion from no-show to show based on previous research done on the effectiveness of text intervention. We believe that the recommendation system would provide more effective outreach, leading to higher conversion rate. This is due to the fact that our recommendation system finds similar appointments where the only difference in the no-show and show result may be if there was intervention (but we cannot know for sure). But for the sake of the evaluation, we will assume a 60% conversion rate, which we examined at 3 cutoffs.

At a .3 similarity rate cutoff, we have 654 patients we can contact: 249 TP, 364 FP, 31 TN, and 10 FN. If we follow our assumptions (with now 60% conversion) and calculation equations from above, our calculations look like this.

|  |  |
| --- | --- |
| **Profit Recovered from Patients Predicted to be No-Shows Correctly** | 249 \* .6 \* ($104 - $0.005) = **$15,536.85** |
| **Costs of Contacting Patients that were Show**  **(Predicted No-Show incorrectly)** | 364\* $0.005 = **$1.82** |
| **Costs of Contacting No-show Patients that still are No-Show** | 249\* .4 \* .005 = **$0.59** |

Based on this cutoff of .3 minimum similarity, We have a final equation of: $15,536.85 - $1.82 - $0.59 - $0.70 = $15,533.743. That is a recovered profit of $15,533.743 over these 2 monthsout of a possible loss of $25,896 (from 249 \* $104), almost 60% ratio.

At a .5 similarity rate cutoff, we have 530 patients we can contact: 198 TP, 296 FP, 27 TN, 9 FN.

|  |  |
| --- | --- |
| **Profit Recovered from Patients Predicted to be No-Shows Correctly** | 198\* .6 \* ($104 - $0.005) = **$12,354.61** |
| **Costs of Contacting Patients that were Show**  **(Predicted No-Show incorrectly)** | 296\* $0.005 = **$1.48** |
| **Costs of Contacting No-show Patients that still are No-Show** | 198\* .4 \* .005 = **$0.40** |

Based on this cutoff of .5, We have a final equation of: $12,354.61 - $1.48 - $0.40 - $0.70 = $12, 345.73. That is a recovered profit of $12, 345.73 over these 2 monthsout of a possible loss of $20,592(from 198 \* $104), about 60% ratio as well.

And finally, at a .7 similarity rate cutoff, we have 296 patients we can contact: 102 TP, 164 FP, 23 TN, and 7 FN.

|  |  |
| --- | --- |
| **Profit Recovered from Patients Predicted to be No-Shows Correctly** | 102\* .6 \* ($104 - $0.005) = **$6, 364.61** |
| **Costs of Contacting Patients that were Show**  **(Predicted No-Show incorrectly)** | 164\* $0.005 = **$0.82** |
| **Costs of Contacting No-show Patients that still are No-Show** | 102\* .4 \* .005 = **$0.20** |

Based on this cutoff of .7, We have a final equation of: $6,364.61 - $0.82 - $0.20 - $0.70 = $6,362.89. That is a recovered profit of $6,362.89. over these 2 monthsout of a possible loss of $10,608(from 102 \* $104), about 60% ratio as well.

With this hypothetical evaluation, we find that there is a better return on our outreach efforts compared to the classification model. However, overall, the recovered profit is less. Why would a clinic choose to use this recommender system over the classification system? Well, again due to HIPAA laws, the clinics must manually text each patient, so there is a large time cost of receptionists’ labor, whose labor may be more useful in other clinical activities. While the ratio of effort is about 60% at all cutoffs, a higher cutoff will help the receptionists text less appointment patients starting from the most likely to convert, and therefore a better expenditure of effort and labor.

To truly evaluate the effectiveness of our recommender system, we will have to collect further data to see how many of the people whom we predicted to be no-show, actually showed up after we sent them the text reminder. Since our classification model has a reliable precision and recall, we can then calculate how many of those who showed up are actually a result of a text reminder.

# 7 Deployment

The deployment comes in two stages. The first stage is a classification model which we will run periodically on all new appointments, with the goal of finding out which appointments will likely be no-show patients. Ideally, this model would be built at the clinic level, as opposed to across all clinics like we did in our model. After that, we will feed these groups of appointments into our recommender system, which will tell us among the group of people we predicted to be no-show patients, how likely they will show up if we send them a text reminder, and the likelihood is the similarity between the predicted no-show patients to those who received text reminders and showed up. The higher the similarity, the higher the likelihood they will show up if texted a reminder.

Since we are not using any sensitive information such as race or nationality, there should not be an ethical problem. Our model is also flexible enough that whenever hospitals decide to collect more information, we can incorporate those additional features into both classification model and recommender system. However, there is possibility for medical discrimination, as there are variables in our model that is reliant on patient medical history. If a medical condition shows a propensity for no-shows to appointments, those patients may be discriminated against as their slots could be reassigned without warning.

One potential risk not necessarily has to do with our model, but with the deployment, is the possibility of patient information leakage. Our proposed plan does not include an encryption of patient related information such as the phone number we will use to contact each patient and send them reminders. Even though there is no patient name related to each appointment, hospitals should still adopt security measures when incorporating our models into their ERP system to prevent any leakage of sensitive information such as patient’s phone number and remain HIPAA compliant.

# 8 Team Member Contributions

Eleanor Lee - Data Understanding, Data Preparation, Powerpoint, Writing

Yifei Ren - Business Understanding, Evaluation, Writing/Editing

Howard Wang – Modeling, Deployment, Powerpoint, Writing

Julie Wang – Data Preparation, Modeling, Evaluation, Writing/Editing

# 9 Bibliography

“#1 Twilio Alternative - Compare Bandwidth vs. Twilio - Save up to 50%.” *The Smarter Twilio Alternative*, Bandwidth, www.bandwidth.com/twilio-alternative/?\_bt=285684301380&\_bk=twilio%2Bpricing%2Bsms&\_bm=e&\_bn=g&kmca=703286323&kmag=32468662250&gclid=Cj0KCQjwybD0BRDyARIsACyS8msPD3yDZiSVxaS8ucB6lEBYUnwkhw5eV5Iw\_z8qThSWo-hxNZffyBoaAomlEALw\_wcB.

Fay, Bill. “How Much A Doctor Visit Will Costs You - Blue Book Prices.” *Debt.org*, 2 Apr. 2019, www.debt.org/medical/doctor-visit-costs/.

Guy, Rebecca, et al. “How Effective Are Short Message Service Reminders at Increasing Clinic Attendance? A Meta-Analysis and Systematic Review.” *US National Library of Medicine*, Blackwell Science, Inc., Apr. 2012, www.ncbi.nlm.nih.gov/pmc/articles/PMC3419880/.

Kent, Jessica. “Predictive Analytics, EHR Data Identify Appointment No-Shows.” *HealthITAnalytics*, Xtelligent Healthcare Media, 14 Aug. 2018, healthitanalytics.com/news/predictive-analytics-ehr-data-identify-appointment-no-shows.

Knowles, Megan. “Johns Hopkins Researchers Develop Tool to Predict Patient No-Shows.” *Becker's Hospital Review*, Becker's Healthcare, 5 Oct. 2018, www.beckershospitalreview.com/patient-experience/johns-hopkins-researchers-develop-tool-to-predict-patient-no-shows.html.

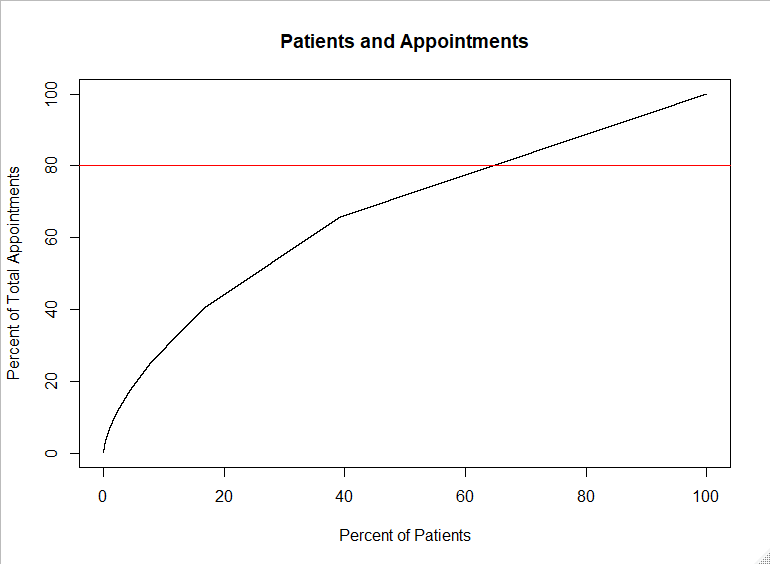
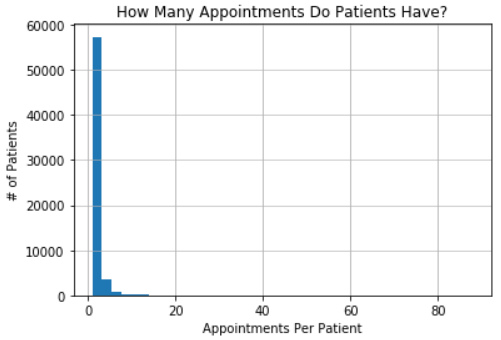
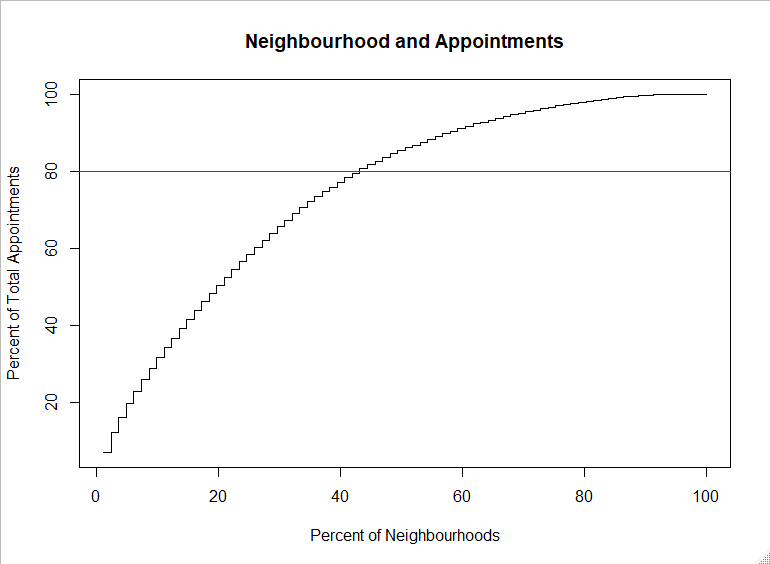
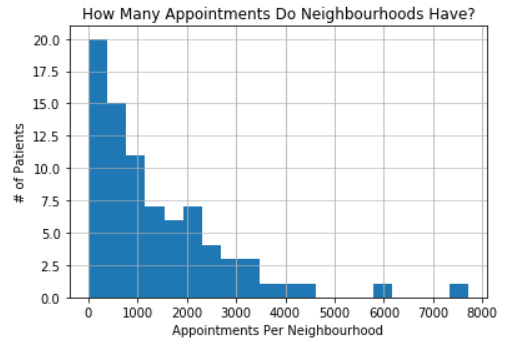
Lin, Chia-Lei, et al. “Text Message Reminders Increase Appointment Adherence in a Pediatric Clinic: A Randomized Controlled Trial.” *NCBI*, US National Library of Medicine, National Institutes of Health, 29 Dec. 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5227159/.

“NHE Fact Sheet.” *CMS.gov*, U.S. Centers for Medicare & Medicaid Services, 24 Mar. 2020, 4:15, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.

Pinsky, Ilana, et al. “Patterns of Alcohol Use among Brazilian Adolescents.” *Brazilian Journal of Psychiatry*, Associação Brasileira De Psiquiatria, 30 Apr. 2010, www.scielo.br/scielo.php?script=sci\_arttext&pid=S1516-44462010000300007.

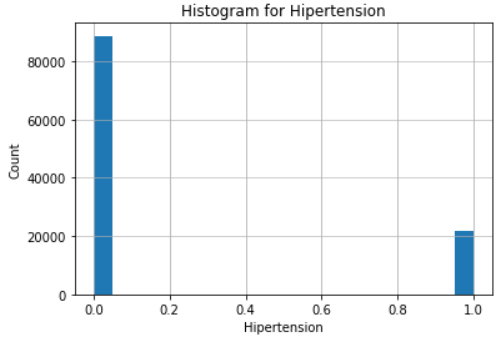
# 10 Appendix/Screenshots

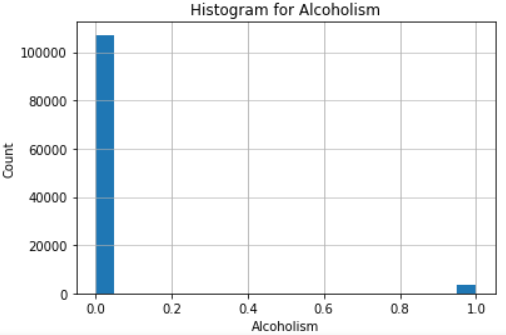
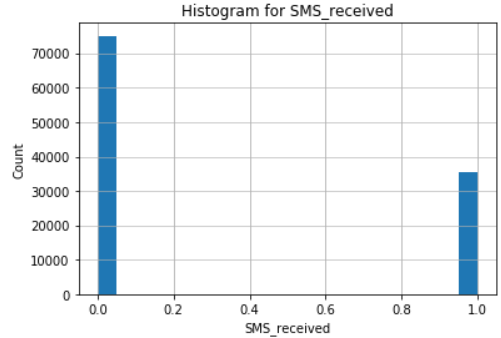
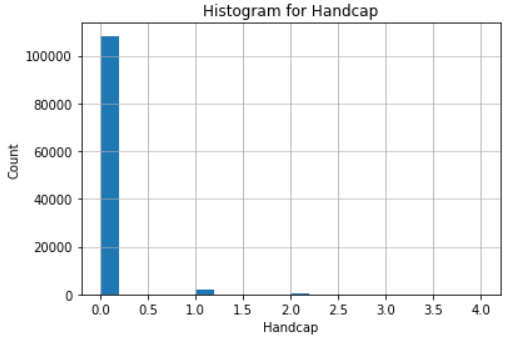
## 10.1 Understanding Appointment Distribution

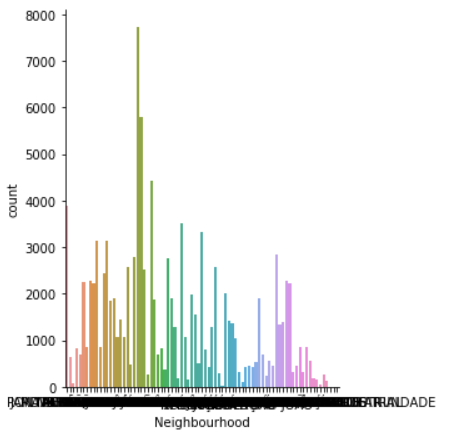


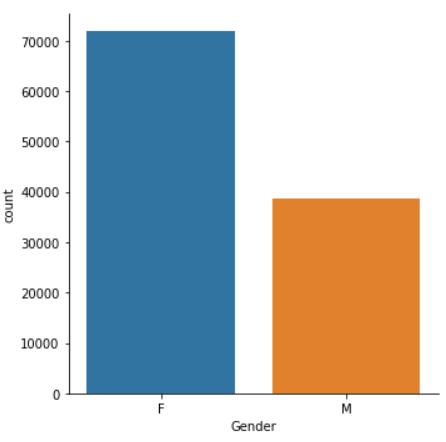
## 10.2 Histograms of Variable Distribution & Correlation

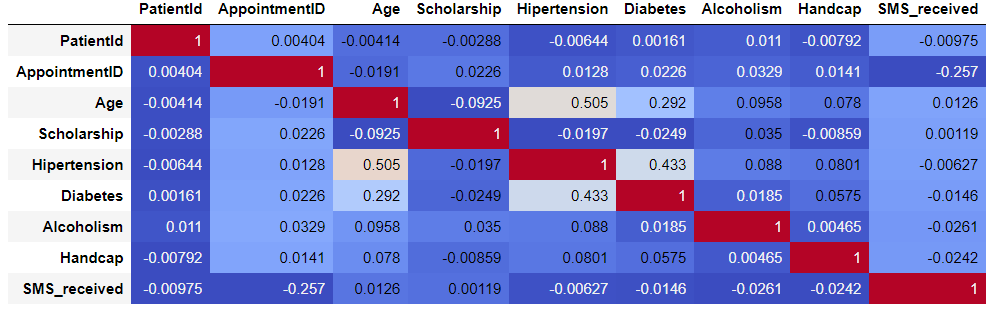
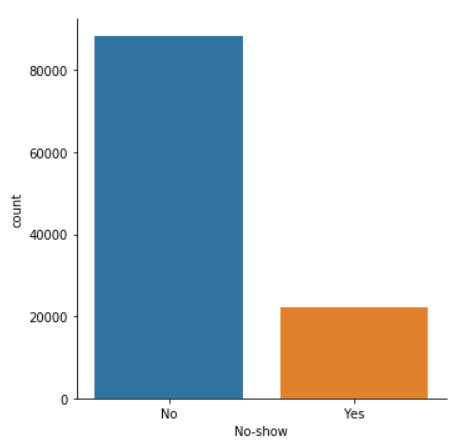
## 





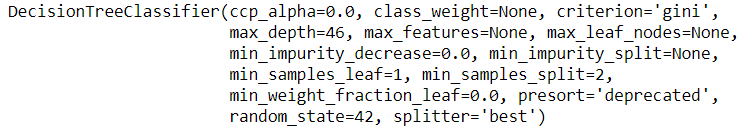


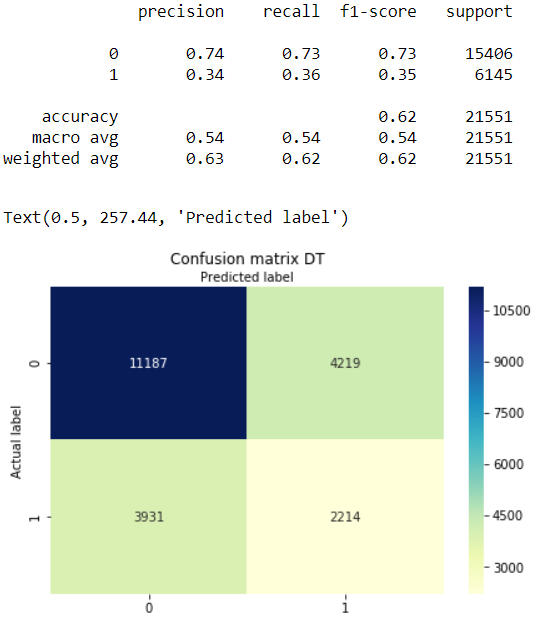


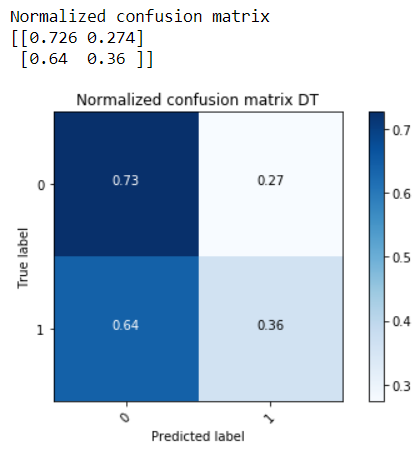
****

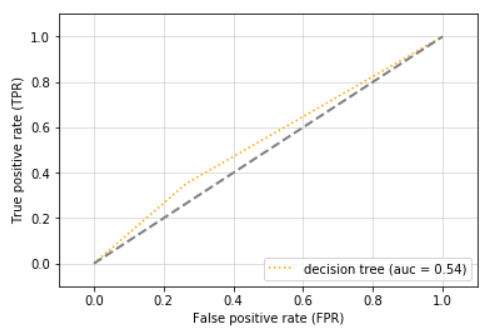
## 10.3 Classification Models

**DT**



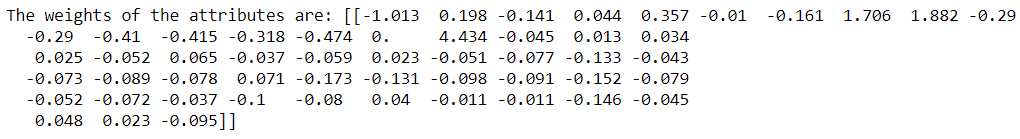


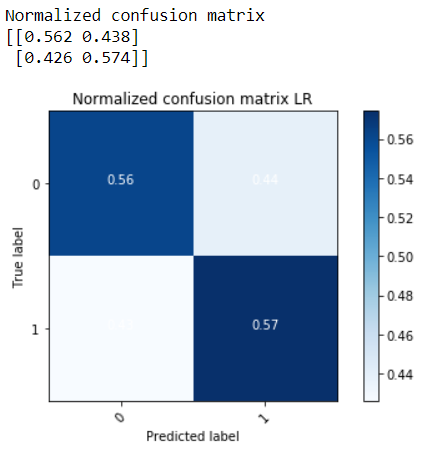
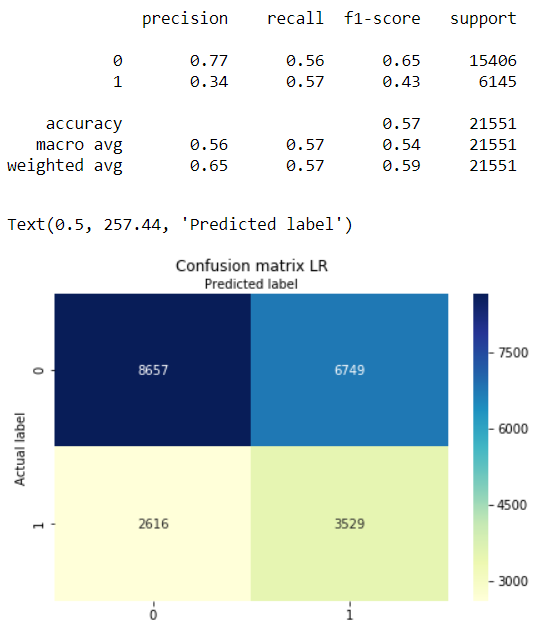


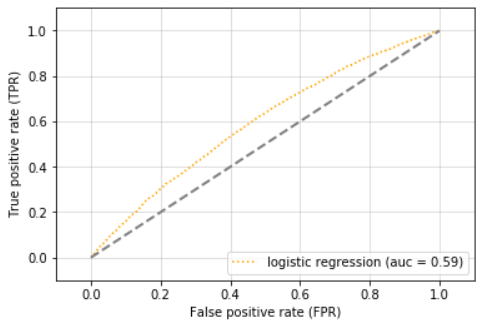


**LR**

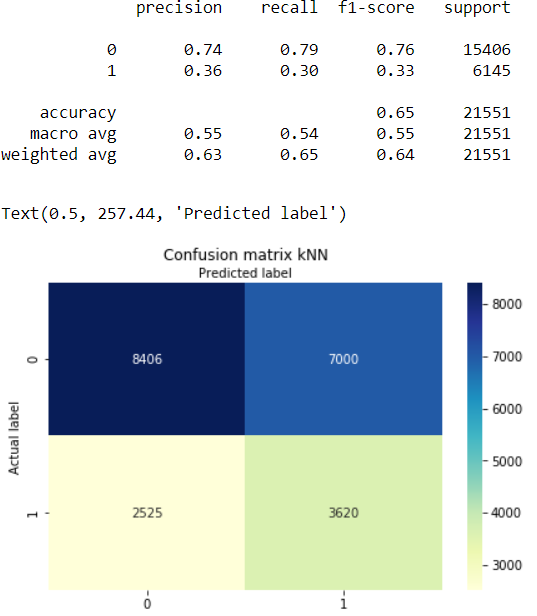
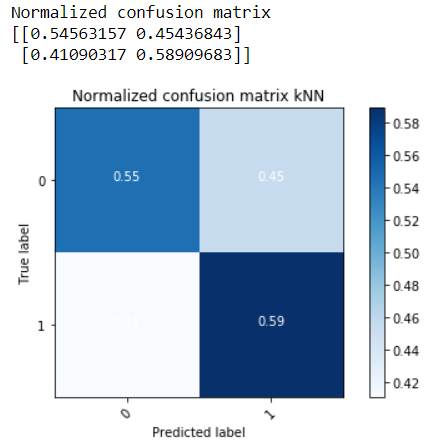
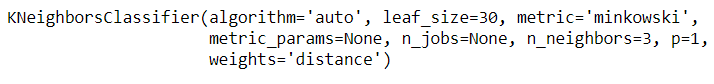


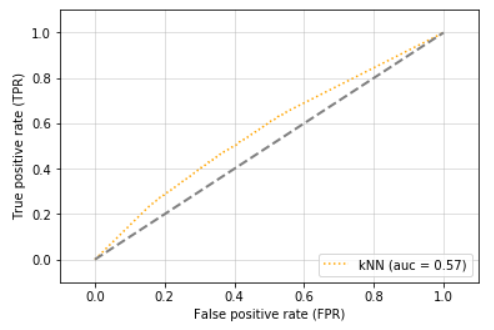




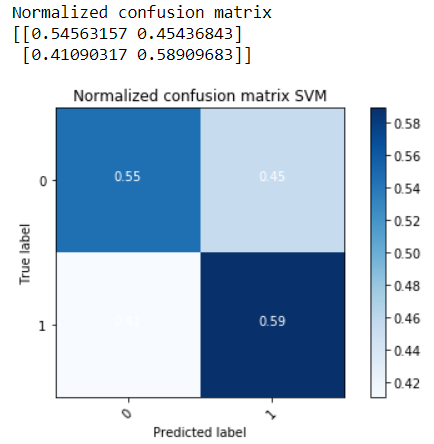
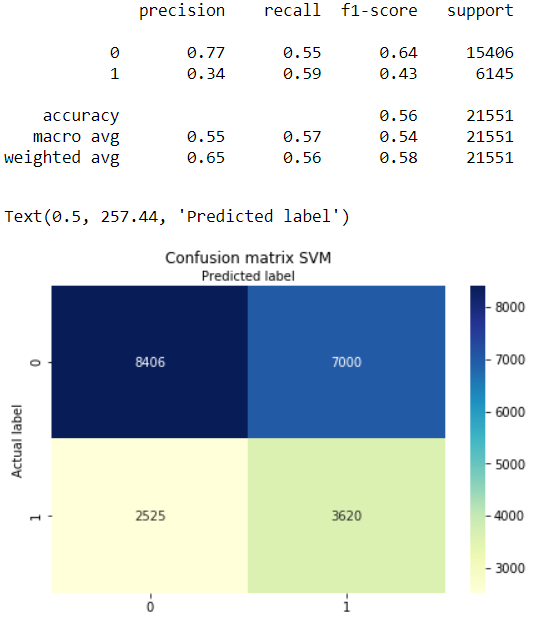


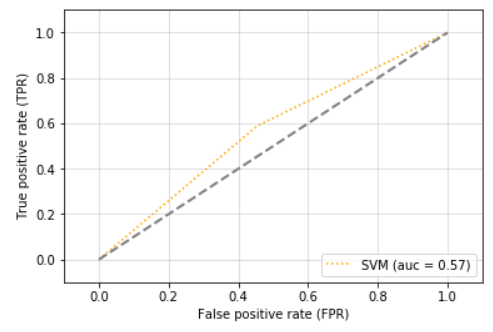
**kNN**



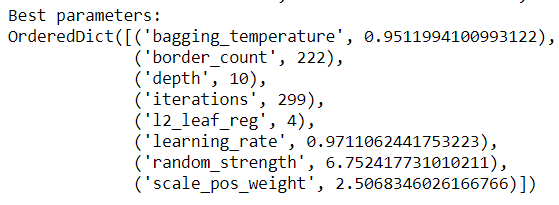


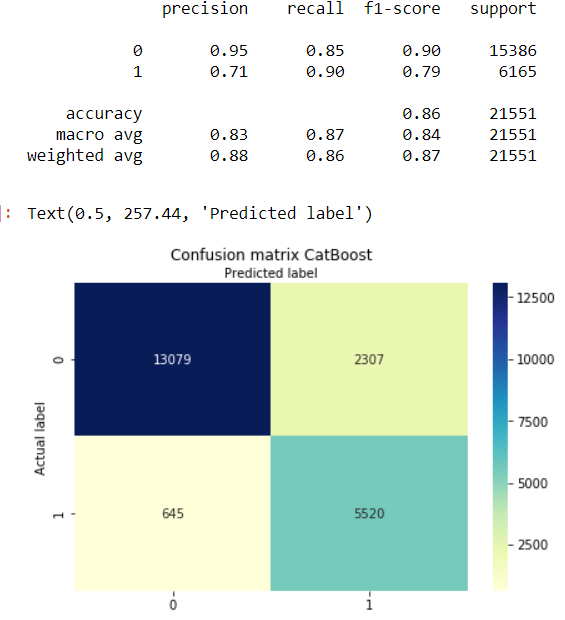
**SVM**

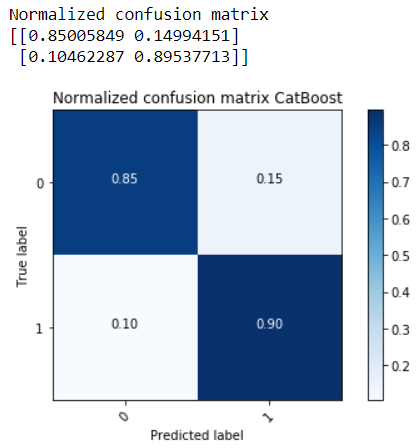


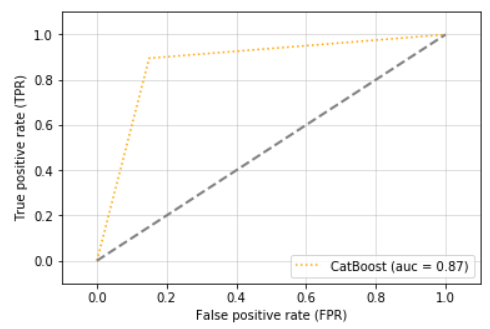


**CatBoost**

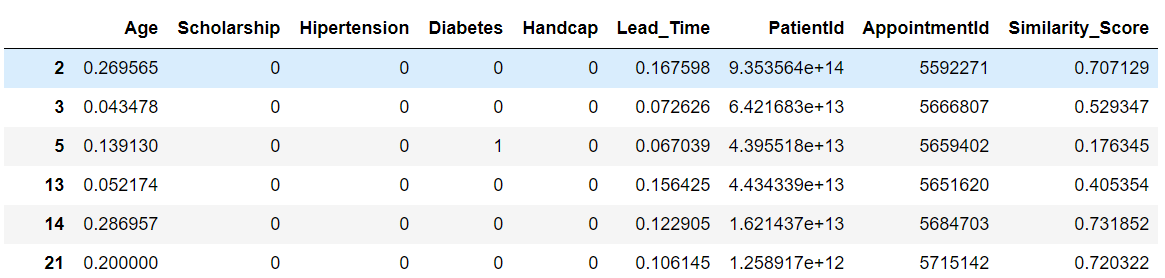
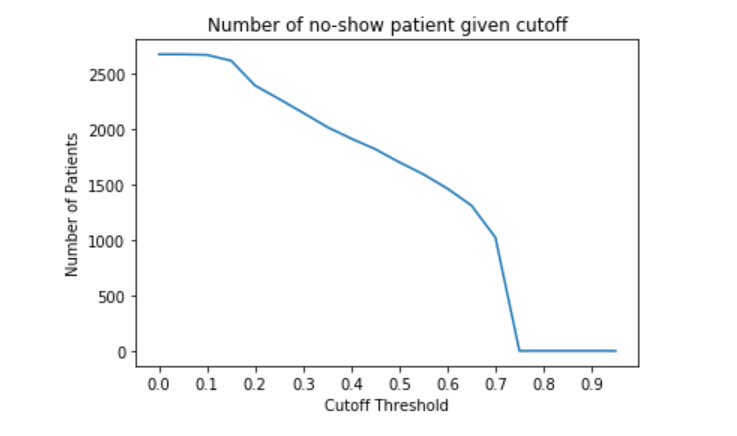
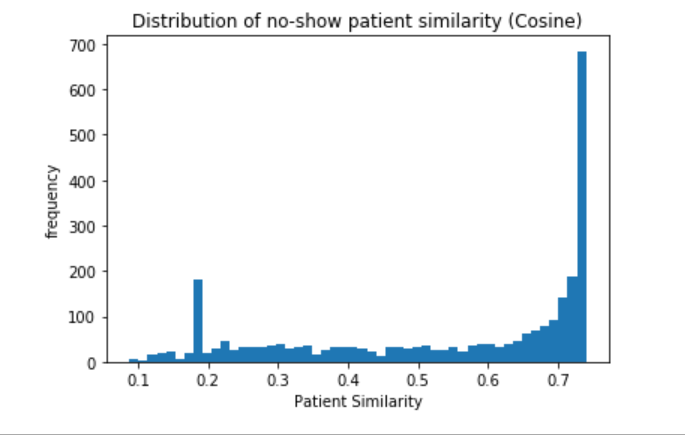








## 10.4 Recommender System



1. “NHE Fact Sheet.” *CMS.gov*, U.S. Centers for Medicare & Medicaid Services, 24 Mar. 2020 [↑](#footnote-ref-1)
2. Fay, Bill. “How Much A Doctor Visit Will Costs You - Blue Book Prices.” *Debt.org*, 2 Apr. 2019 [↑](#footnote-ref-2)
3. Knowles, Megan. “Johns Hopkins Researchers Develop Tool to Predict Patient No-Shows.” *Becker's Hospital Review*, Becker's Healthcare, 5 Oct. 2018 [↑](#footnote-ref-3)
4. Kent, Jessica. “Predictive Analytics, EHR Data Identify Appointment No-Shows.” *HealthITAnalytics*, Xtelligent Healthcare Media, 14 Aug. 2018 [↑](#footnote-ref-4)
5. Lin, Chia-Lei, et al. “Text Message Reminders Increase Appointment Adherence in a Pediatric Clinic: A Randomized Controlled Trial.” *NCBI*, US National Library of Medicine [↑](#footnote-ref-5)
6. Pinsky, Ilana, et al. “Patterns of Alcohol Use among Brazilian Adolescents.” *Brazilian Journal of Psychiatry*, Associação Brasileira De Psiquiatria, 30 Apr. 2010 [↑](#footnote-ref-6)