

New York University Stern School of Business

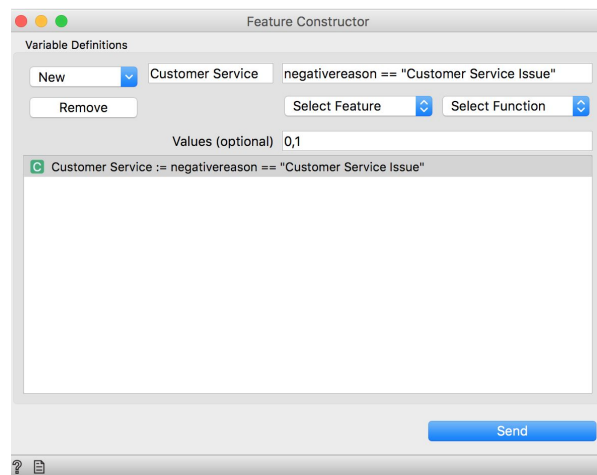
# Group Assignment #3

Kelly Leong, Lucy Lian, Teddy Le Nguyen, Nancy Toussie  
Data Mining For Business  
Professor Fernandez  
Spring 2020

The target variable is whether or not a tweet is related to customer service issues.

We want to identify which tweets are related to customer service centers to figure out why customers are dissatisfied with TAA's customer service.

We loaded the 4000 tweets we are studying into Corpus. We used Feature Constructor to create a new binary target variable called "Customer Service" which takes the value of 1 if "negativereason" is "Customer Service Issue" and 0 if else. We used Select Columns to assign "customerservice" as the target variable.



We used Select Columns to assign the "tweet\_id", "text", "airline\_sentiment", and "negativereason" as the meta attributes. We set our newly created variable "Customer Service" as the target variable. We use Select Rows to filter the data on our target variable "Customer Service". When "Customer Service" = 1, this is all the negative tweets that correspond to a customer service issue. When "Customer Service" = 0, this is all positive tweets and any negative tweets that don't correspond to customer service. We viewed three examples of positive tweets and three examples of negative tweets in the Corpus Viewer.

Select Columns

Available Variables

Filter

Features

Filter

Up

>

Down

Up

>

Down

Target Variable

Up

>

Down

Customer Service

Meta Attributes

Up

>

Down

tweet\_id

airline\_sentiment

negativereason

text

Reset

☒ Send Automatically

Corpus Viewer

Info

Documents: 3178

Preprocessed: True

◦ Tokens: 71518

◦ Types: 7373

POS tagged: False

N-grams range: 1-1

Matching: 3178/3178

Search features

Customer Service

tweet\_id

airline\_sentiment

negativereason

text

Display features

Customer Service

tweet\_id

airline\_sentiment

negativereason

text

☐ Show Tokens & Tags

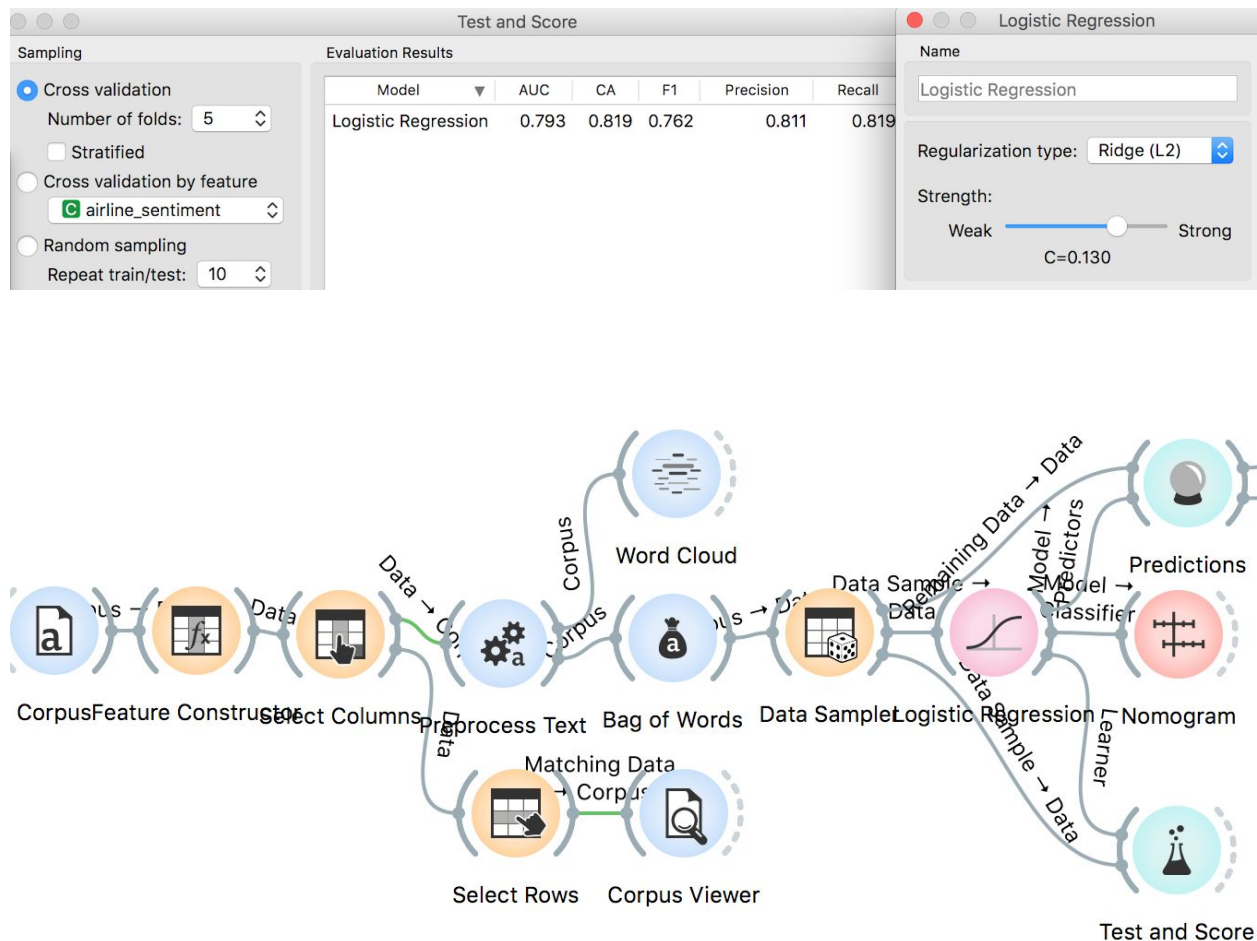
☒ Auto send is on

RegExp Filter:

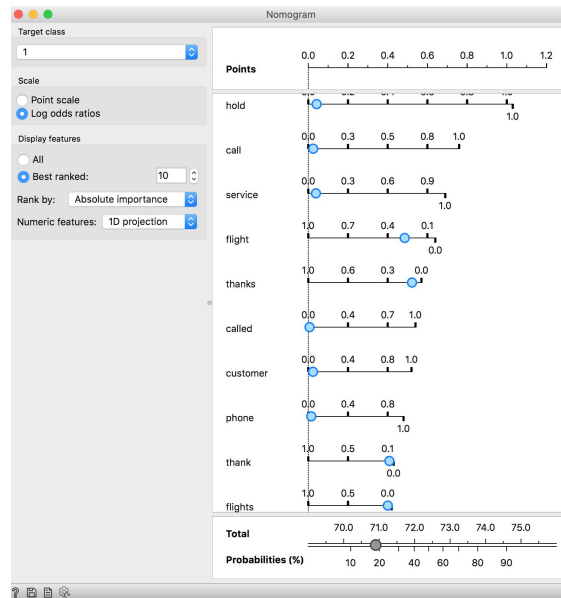
1	@TAA site errored out at last step of changing award. Now can't even pull ...	<div>Customer Service: 0</div>
2	@TAA Port-Au-Prince, Haiti for continued food distribution, medical suppli...	<div>tweet_id: 5.7e+17</div>
3	@TAA Having an issue finding a missing item on a plane. Can you help me f...	<div>airline_sentiment: positive</div>
4	@TAA what is the problem with your flights to ny today	<div>negativereason: ?</div>
5	@TAA We have been stuck in SJU for several hours and no one is answerin...	<div>text: @TAA thanks! Very excit-ed to see it :D</div>
6	@TAA She seems a little preoccupied - that's why I'm bringing it to your att...	<div>Customer Service: 0</div>
7	@TAA flight #703 Philly -&gt; Mia? Bwahahaha! Delayed 5+ hrs, were in th...	<div>tweet_id: 5.7e+17</div>
8	@TAA I'll have to drop by next time I visit!	<div>airline_sentiment: positive</div>
9	@TAA can you tell me why flight 1542 from dfw to phl was Cancelled Flight...	<div>negativereason: ?</div>
10	@TAA thanks! Very excited to see it :D	<div>text: "@TAA: @2littlebirds Beautiful shot.. Thanks for sharing. Using #FlyFi to post? :)" Your welcome! Not on this flight. It was a quickie;)</div>
11	@TAA any chance of adding LAX-&gt;JFK direct any time in the future?	<div>Customer Service: 0</div>
12	@TAA - I did that as soon as I realized. Will they try to figure out who is ste...	<div>tweet_id: 5.7e+17</div>
13	"@TAA: @2littlebirds Beautiful shot.. Thanks for sharing. Using #FlyFi to p...	<div>airline_sentiment: positive</div>
14	@TAA is probably losing a family of world travelers as customers due to br...	<div>negativereason: ?</div>
15	@TAA 40+ minutes waiting at the gate for the broken jetway. Different gate...	<div>text: @TAA would also like say kind move on adding the points !</div>
16	@TAA DCA-&gt;HPN, u r prof hostage takers.We stood on bus 2 plane for ...	
17	@TAA would also like say kind move on adding the points !	
18	@TAA MR, she's on her way now, but thought id detail the extravaganza fo...	



different complexity parameters. We used AUC because it corresponds to the area under the ROC curve; the ROC curve summarizes the performance of a binary classification model on the positive class, and therefore, AUC is a summary statistic that can tell us how well a model performs. We applied Lasso and Ridge regularization with three complexity control parameters at  $C=1$ ,  $C=0.130$ , and  $C=0.08$  (due to the fact that our Orange was taking extremely long). The combination that best maxed AUC was Ridge (L2) regularization at a strength  $C=0.130$ . The resulting AUC has the value of 0.793, indicating that the model is much more predictive than a random model.



Since we are particularly interested in identifying negative tweets that are related to customer service issues, the words that caught our attention are words that are related to the possible reasons behind negative customer service issues sentiment, such as “hold”, “call”, “called”, and “phone”. These words hint at the reasons why customers may be upset with TAA’s customer service -- particularly issues having to do with calling, answering, and holding customers on the phone. It is also useful to know that the words “customer” and “service” are explicitly present in many of the tweets related to customer service issues.



We used Bag of Words to choose how we want our model to represent words. We found that binary variables performed better on AUC with a value of 0.793 versus an AUC of 0.792 for count. This slight difference of .001 might be a coincidence in the sample and cannot be generalized, therefore we cannot conclude binary is the best representation based on this alone. Binary's slightly superior performance could be due to the fact that tweets are very short, and people are less likely to repeat the same words than if they were writing through another medium. Thus, we decided to proceed with a binary representation under the assumption the presence, instead of count of words, are enough to be predictive of our target variable.

Test and Score							Bag of Words			
Evaluation Results							Options			
Model	AUC	CA	F1	Precision	Recall		Term Frequency:			
Logistic Regression	0.793	0.819	0.762	0.811	0.819		Binary			
							Document Frequency:	(None)		
							Regularization:	(None)		
							<input checked="" type="checkbox"/>	Commit Automatically		

Test and Score							Bag of Words			
Evaluation Results							Options			
Model	AUC	CA	F1	Precision	Recall		Term Frequency:			
Logistic Regression	0.792	0.823	0.769	0.819	0.823		Count			
							Document Frequency:	(None)		
							Regularization:	(None)		
							<input checked="" type="checkbox"/>	Commit Automatically		

We are interested in viewing the incorrectly classified tweets to evaluate our model's performance, so we linked the remaining data (test data) to the Predictions. A false positive means that the model predicted



the tweet was a negative tweet about customer service ( $LR > 0.5$ ) when it was actually either a positive tweet or a negative tweet not involving customer service (Customer Service = 0). A false negative means that the model predicted the tweet was a positive tweet or a negative tweet not involving customer service ( $LR < 0.5$ ), when it was actually a negative tweet about customer service (Customer Service = 1). We added these conditions to Select Rows to be viewed on Corpus Viewer with the results included below.

Upon viewing the false negative tweets, we noticed that the negative words appearing in these tweets were not frequent across all other tweets. Therefore, the error could lie in that the model was not given tweets regarding these types of negative incidents (e.g. “nasty”, “poor”, "rudeness"), since these types of words were not frequent in the Nomogram. This lack of exposure thereby decreases the model's predictiveness on those words. To rectify this, we could increase the data that we used for the training set, as we had only used 25% of the data to build the model. The model could learn more words that are predictive of negative sentiment towards customer service, and therefore lessen the number of false negatives.

The screenshot shows the 'Corpus Viewer (1)' application. On the left, there are panels for 'Info' (Documents: 523, Tokens: 5078, Types: 6948, POS tagged: False, N-grams range: 1-1, Matching: 523/523) and 'Search features' (Customer Service, tweet\_id, airline\_sentiment, negativereason, text, Logistic Regression, Logistic Regression (0), Logistic Regression (1)). Below these are 'Display features' and a checkbox for 'Auto send is on'. The main area displays a list of tweets with their corresponding Customer Service status and Logistic Regression scores.

Index	Tweet Text	Customer Service	Logistic Regression (0)	Logistic Regression (1)
1	@TAA Boston is all self ser...	1		
2	@TAA was that English? i'll...			
3	@TAA you would rather a ...			
4	@TAA congrats on treating...			
5	@TAA Thanks for your can...			
6	@TAA this customer servic...			
7	@TAA Just saying the trut...			
8	@TAA requested a call bac...			
9	Is rudeness and poor cust...			
10	@TAA I need assistance wi...			
11	@TAA your team was just t...			
12	@TAA yes I did but no resp...			
13	@TAA			
14	@TAA I had the worst cust...			
15	@TAA if you are going to ...			
16	@TAA poor showing today...			
17	@TAA I've been on hold fo...			
18	@TAA I spoke with a repre...			
19	@TAA your customer servi...			
20	@TAA your mobile site is b...			
21	@TAA I have lots of flights ...			
22	@TAA @TAA Is there a wa...			

Upon viewing the false positive tweets, we noticed that these tweets contained words that the model had learned to be highly predictive of customer service issues (i.e. "call", "hold"). All of the tweets were negative, but not about customer service. This might be due to the fact that words the model identified as highly predictive are in fact too general to narrow down to specifically a customer service issue -- the customer had called customer service, but why? To rectify this, we could amend the stopwords file to include tokens on the nomogram that we find too general, until the model's predictiveness decreases (lowering our AUC).

Corpus Viewer (1)		
Info		
Documents: 9		
Preprocessed: True		
Tokens: 112		
Types: 6948		
POS tagged: False		
N-grams range: 1-1		
Matching: 9/9		
Search features		
<input checked="" type="checkbox"/> Customer Service <input checked="" type="checkbox"/> tweet_id <input checked="" type="checkbox"/> airline_sentiment <input checked="" type="checkbox"/> negativereason <input checked="" type="checkbox"/> text <input checked="" type="checkbox"/> Logistic Regression <input checked="" type="checkbox"/> Logistic Regression (0) <input checked="" type="checkbox"/> Logistic Regression (1)		
Display features		
<input checked="" type="checkbox"/> Customer Service <input checked="" type="checkbox"/> tweet_id <input checked="" type="checkbox"/> airline_sentiment <input checked="" type="checkbox"/> negativereason <input checked="" type="checkbox"/> text <input checked="" type="checkbox"/> Logistic Regression <input checked="" type="checkbox"/> Logistic Regression (0) <input checked="" type="checkbox"/> Logistic Regression (1)		
<input type="checkbox"/> Show Tokens & Tags <input checked="" type="checkbox"/> Auto send is on		
RegExp Filter:		
1 @TAA Flight's Cancelled Flig...	Customer Service:	0
2 @TAA my flight has been del...	negativereason:	Cancelled Flight
3 @TAA You don't have those a...	text:	@TAA Flight's Cancelled Flightied. Website says to call phone number.Phone says to check online. How am I supposed to get some help?
4 @TAA found great prices thr...	Logistic Regression:	1
5 @TAA no i use to skyscanner...	Logistic Regression (0):	0.34176
6 @TAA Lost bag process is br...	Logistic Regression (1):	0.65824
7 @TAA when I called I was tol...	Customer Service:	0
8 @TAA Need to talk to a perso...	negativereason:	Flight Booking Problems
9 @TAA On hold waiting to ma...	text:	@TAA found great prices thru US Air for honeymoon. Website wont work and been on hold for more than a hour now after being told 10 min
	Logistic Regression:	1
	Logistic Regression (0):	0.448795
	Logistic Regression (1):	0.551205
	Customer Service:	0
	negativereason:	Lost Luggage
	text:	@TAA when I called I was told my bag had made it to PHL, but still has not been delivered or any call from the delivery service
	Logistic Regression:	1
	Logistic Regression (0):	0.297245
	Logistic Regression (1):	0.702755

As mentioned, we chose to evaluate our model with AUC because we were not given expected costs or benefits. On the test data, the model produced an AUC of .807, indicating good predictiveness. We examined this further by exporting the “Customer Service,” “text,” and “Logistic Regression (1)” columns to a CSV, using Select Columns and Save Data. We looked at the tweets in descending order, the top tweet being with the highest probability of a negative tweet about customer service (highest numeric value for LR). Our ultimate goal is to send the top 20 out of 1000 tweets, which should all definitively be negative customer service issue tweets, directly to management. Keeping this ratio, we looked at the top 60 tweets out of the 3000 test tweets we had employed our model on. In these 60 tweets, there were 2 incorrectly classified false positive tweets.

To report the results to general management, we would explain that the model made very little errors, only 2 out of 60 (3.33%), when we tested it with significant amount of test data and therefore, would likely do a good job performing on 1000 tweets each day to find the top 20 most likely to be customer service issue related. Despite these errors, the model is useful because even a brief glance at the tweets can tell us which problems are the most frequent (in the case, hold times).



Customer	Logistic R	text
1	0.93025009	@TAA I have been on hold w/customer service line for 68 minutes. This after I was on phone with an agent for 35 min b/f call dropped
1	0.8424286	@TAA So today I call into Customer Service & speak with Wendell Holton in the Dallas office & I am told there is nothing he can do
1	0.80555268	@TAA I've been on hold for 90 minutes. This is terrible customer service. You messed up my reservation! <a href="http://t.co/oYlgeao7Y8">http://t.co/oYlgeao7Y8</a>
1	0.7945007	@TAA I've been on hold with customer service for over an hour. Can you help?!
1	0.76878018	@TAA that's absolutely horrible customer service. The person supposed to call me can't immediately call back when disconnected???
1	0.73527737	@TAA that's unacceptable. They should allow me to wait on hold, or take my number and call me themselves.
1	0.7344884	@TAA Any plans to implement a call-back system on your reFlight Booking Problems line? Been on hold for over an hour now...
1	0.73079674	@TAA when trying to check-in online, it says to call...now I've been on hold for 2 hours...what to do?
1	0.72653186	@TAA But every time I call I'm on hold for 45 min and then get disconnected. Any way someone can help?
1	0.72129931	@TAA Your customer service line keeps telling me call back Late Flightr and disconnects. How can I talk to someone about my flight?
1	0.72067885	@TAA I spent an hour on the phone with customer service only to find out that they can't help. Need help with family pooling asap.
1	0.71849737	@TAA Collectively since Friday I've been on hold with customer service for over 3 hrs. My issue is not resolved... so frustrated!
1	0.71572848	@TAA @TAA as soon as I call customer service to speak to a representative, I am disconnected
1	0.70461492	@TAA How do I check? Reservation for Joe Watson and Kelsey Jennings. We were on hold for 2 hours. Waiting for call back now. ETA?
0	0.70275072	@TAA when I called I was told my bag had made it to PHL, but still has not been delivered or any call from the delivery service
1	0.70243033	@TAA I waited on hold for two hours, only to have my call. Really unreliable.
1	0.6647893	@TAA so upset with customer service. I have a simple question and the phone system has disconnected me all day today and yesterday.
1	0.66001032	@TAA why does your customer service line say "we are experiencing high call volume" and then spontaneously hangup on the person?
0	0.65824008	@TAA Flight's Cancelled Flightled. Website says to call phone number.Phone says to check online. How am I supposed to get some help?
1	0.6579272	@TAA how is it possible that you don't have a call back service? An automated message telling us to call back Late Flightr is so unhelpful.
1	0.65329862	@TAA been on the phone for over an hour with customer service and they can't figure it out! awesome service... NOT!
1	0.65258528	@TAA customer service (if you can call it that) refunded my money
1	0.65067471	@TAA FYI twice now I've tried the "we'll call you when someone is free" thing. Both times it called, I picked up, then it hung up.
1	0.64759363	@TAA Can you help me update KTN on the profile? Ticket book via AA & that info didn't get pass on. I've been on hold forever
1	0.64614965	@TAA little help to us! My husband has been on hold for an hour for something that take 5 minutes to complete.
1	0.64525849	@TAA please send me a number to call to talk to a person and not be put on hold
1	0.64234244	@TAA received an email requesting I call about a res. but I keep getting kicked off of your phone system. Help?
1	0.63345289	@TAA my request has nothing to do with rescheduling. But customer service has such a bad attitude today. #badservice
1	0.62744512	@TAA is struggling. had me on hold for 25 minutes, kicked me back to the start, then said they couldnt handle the amount of calls
1	0.62668875	@TAA Anyone answering phone this morning at 800IFLYSWA? On hold 51 minutes and counting...
1	0.6237224	@TAA I cant get anyone on the phone to help with award travel.Purchased extra miles and 10 minutes Late Flightr the miles needed was raised.
1	0.62125658	@TAA I'm going to need you to answer your phones. Being on hold for an hour isn't okay.
1	0.62125658	@TAA I'm going to need you to answer your phones. Being on hold for an hour isn't okay.
1	0.62066003	@TAA If it was so important, why did I wait on hold and then get hung up on by your computer? #disappointed
1	0.61829454	@TAA my wife has now been on hold for 90 minutes....I get weather is bad, but this is insane. #customerservice #customerloyalty
1	0.61351172	@TAA any effort to fix your points system? 11 days and a call to customer service and still no account update. Chasing points is no fun
1	0.60346528	@TAA You are a F&KING joke. Your customer service is woeful. If your staff tell a passenger that they will do something do it!
1	0.60030689	@TAA I've been on hold for 40 minutes just to get a receipt. I know you guys are busy but I think it's time to higher more staff!
1	0.59937037	@TAA again I ask, who can I call to get this fixed? Tweeting me BS questions to stall only makes things worse. Phone number please.
1	0.59842242	@TAA I need to register a service dog for a first class ticket from SFO &gt; Dulles. The phone queue is an hour or longer. Pls advise
1	0.59824763	@TAA you can't control the weather but you can control customer service. Without luggage for 24 hours & still yet to speak to a human.
1	0.59505922	@TAA help! I've been on hold for almost two hours. Trying 2 get home.
1	0.59485305	@TAA-rebooked to OMA - 180 miles from my destination. Spotty customer service. I get staff stressors but come on, this is your business.
1	0.59426255	@TAA have been on hold for 58 minutes - need help USAIR - why won't you help?
1	0.59426255	@TAA you're customer service is unbelievably bad. Abysmal. I've been on the phone for well over 2hrs tonight. Hung up on by yr ppl 3x.
1	0.5912995	@TAA - I wouldn't need to wait on hold for 6+ hours if your website worked correctly - please respond!
1	0.59082688	@TAA tried twice today on hold for 30 min each time. I have things to do so can't live on hold dealing w/ your customer serv failures
1	0.5907946	@TAA we showed up to our flight several hours early. Tried to call & was on hold 15 min b/f call disconnected.
1	0.58935366	@TAA @SouljaCoy what is AA going to do to fix their utterly embarrassing customer service? You won't even answer the dang phone!
1	0.58690383	@TAA she sent you an email 2 customer service the night it happened ur response was basically 2 bad sorry how should she contact u.
1	0.58650857	@TAA would someone please DM me the customer relations number. The website has only customer service.
1	0.58200962	@TAA awaiting my return phone call, just would prefer to use your online self-service option :
1	0.57941994	@TAA how long does it take to get a response from your customer service about a complaint done through email?
1	0.57807168	@TAA I will be calling someone on Monday in customer relations. Very disappointed in how I was treated by customer service!
1	0.57611081	@TAA there really should not be guidance to call via phone if you are that back logged. I could have done it online in 3 minutes.
1	0.57601586	@TAA so much for being here to help. Very disappointed in your customer service.
1	0.57496624	@TAA I do not want to deal with your customer service agents with no practical knowledge.Provide me with a direct contact.
1	0.57324559	@TAA It says to call. Before connecting, get song, dance about weather. Weather bad 3 days? Called for 2 days before. #wasteoftime
1	0.57204964	@TAA "please call Late Flightr" is not acceptable - but neither is being on hold for 5+ hrs #onholdwith
1	0.56720791	@TAA is there a number to call to modify reservation. Please help. Very frustrating with american and us air merger

Despite the errors mentioned prior, the model is useful because even a brief glance at the tweets can tell us which problems are the most frequent. When we use Predictions to view the top 10 tweets with the highest probability of being negative about customer service (highest LR values), it is obvious that the main customer dissatisfaction with TAA's customer service department is the long hold times.

