Improving Customer Service in Finance using Machine Learning

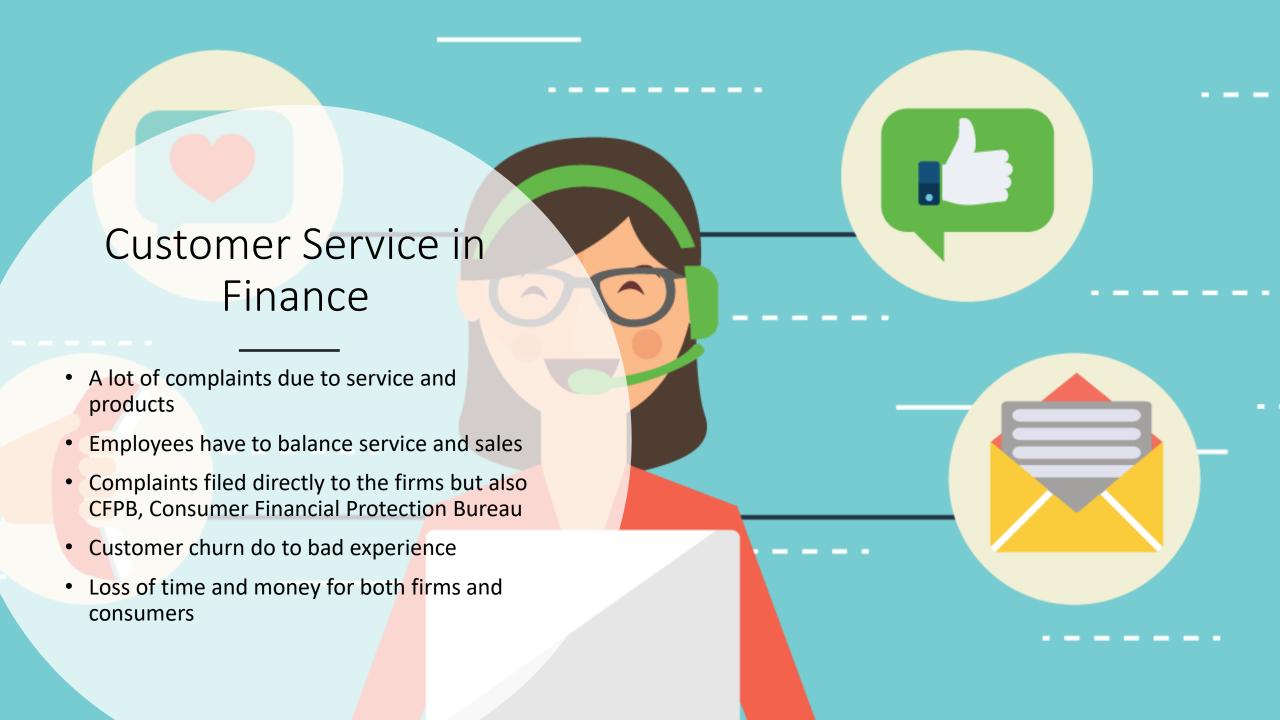
Justin Tsao

Data Science Capstone Project, June 2019



Financial Firms

- Include banks, mortgage lenders, brokerage firms, credit companies
- Products: mortgages, loans, credit cards, bank accounts, financial services
- Very sales and marketing focused
- Handle sensitive consumer information
- Requires consumer to put their trust into the firms hands



Complaints

- For the most part, many complaints are handled appropriately by firms
- Employees trained in customer service
- Some complaints are not handled so well
- Filed online, in person, over the phone
- Stressful and emotional process



Problems to solve

How can we predict difficult to handle consumer complaints in finance?

What do these difficult complaints look like?

If we can predict, what can be done to improve customer service?

Who can benefit?

Consumers





Financial Firms

































Benefits



Save time and money for consumer and firms



Prevent customer churn



Being able to predict right away which complaints need special handling



Leverage results to improve internal customer service policies and procedures



Prevent mishandling of consumer information

The Data

- The data used to answer the previous question comes directly from the CFPB
- Contains over 1,200,000 complaints as of 5/20/2019
- Filed directly by consumers against various firms
- Data can be found https://data.world/cfpb/consumer-complaints



Data Overview

Contains 18 columns

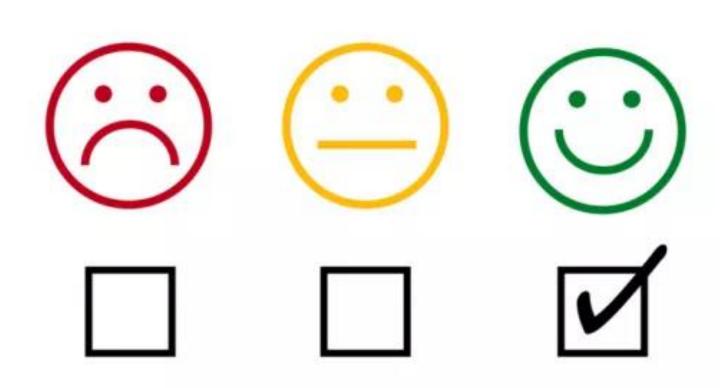
1,284,185 rows

Contains a feature for written complaint called 'consumer_complaint_narrative'

Complaints are those filed to CFPB and reflects just a sample of all complaints filed to financial firms since 2012

How do we predict the difficult complaints?

- Data contains a column called 'timely_response'
- A timely response will be our indicator of whether our complaint was handled in an 'appropriate' manner or not
- 'timely_response' will be the predictor label for machine learning
- True or False

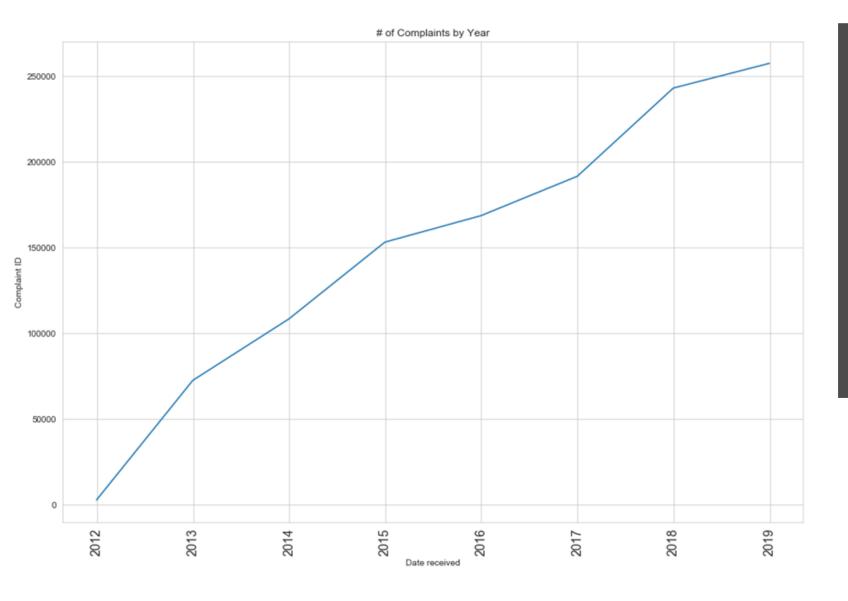


Defining an Appropriate Response

- Handled in time as defined by CPFB 15 days
- Has value of True in data for 'timely_response'
- True assumes complaint handled as efficiently as it could
- Not responding in time possibly means complaint was not resolved, difficult to handle, or maybe ignored due to lack of concern



Exploratory Data Analysis



Complaints by year

- 2013 to 2014 up 149.53%
- 2014 to 2015 up 141.42%
- 2015 to 2016 up 110.09%
- 2016 to 2017 up 113.64%
- 2017 to 2018 up 126.90%
- 2018 to 2019 up 105.93%

Complaints By Year

- Number of complaints filed to CFPB at least doubles each year
- Increased partly due to ease of reporting and adoption of technology
- Worse yet, we can infer customer service is falling or that consumers are increasingly unhappy



of Complaints by Product

Top complaints based on chart

Mortgage: 278249

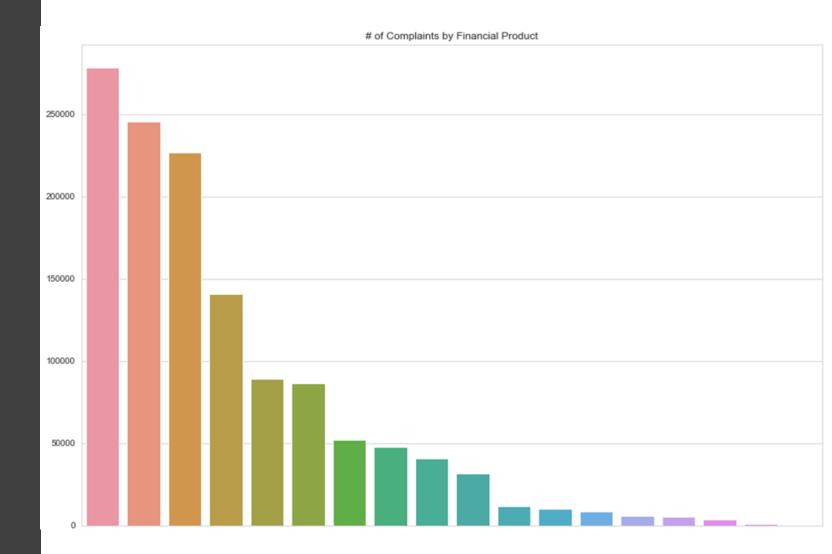
Debt collection: 245218

Credit Based Services: 226781

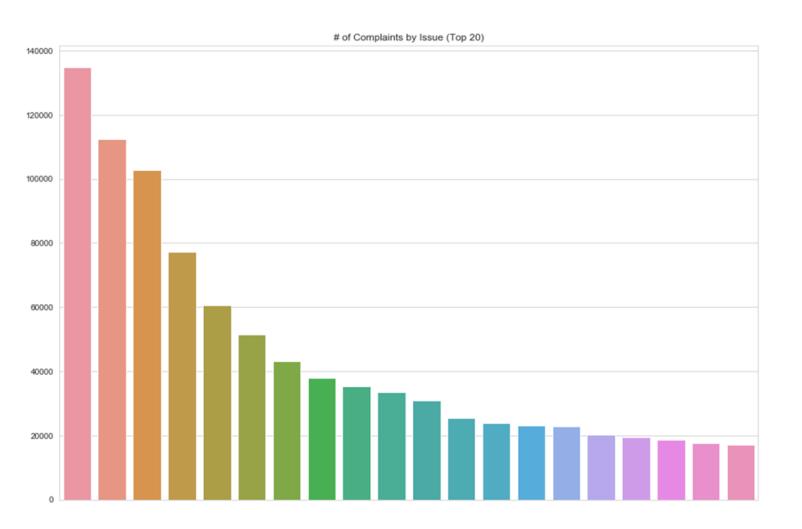
Credit reporting: 140432

Credit card: 89190

Bank account or service: 86206



of Complaints by Issue



Incorrect information on your report: 134809 Loan modification, collection, foreclosure: 112311 Incorrect information on credit report: 102686 Loan servicing, payments, escrow account: 77333 Cont'd attempts collect debt not owed: 60687 Problem with a credit reporting company's investigation into an existing problem: 51498 Attempts to collect debt not owed: 43181 Account opening, closing, or management: 37961 Communication tactics: 35449 Improper use of your report: 33441 Disclosure verification of debt: 30800 Managing an account: 25535 Written notification about debt: 23766 Trouble during payment process: 23188 Deposits and withdrawals: 22851 False statements or representation: 20278 Managing the loan or lease: 19438 Struggling to pay mortgage: 18682 Dealing with my lender or servicer: 17630 Application, originator, mortgage broker: 17229

Complaint Issues



Many issues seem to stem from how customer information is handled and reported



Red flag that employees need to be trained to handle consumer information



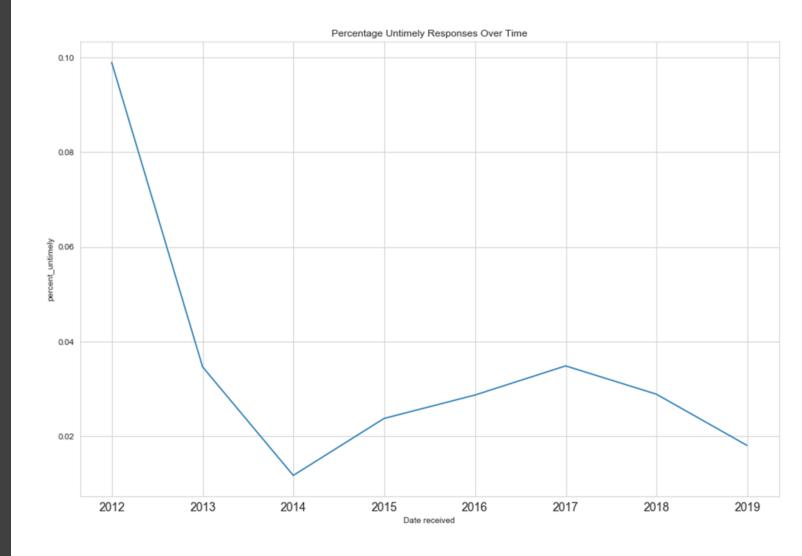
Could be that employees are in a rush to get sales and getting clients in and out of the seat as fast as possible



Entering wrong information can cause many problems for both parties

Percentage of Untimely Responses

- Based on the CFPB's data untimely responses have been decreasing since 2017
- Even a small percentage of complaints not appropriately responded to can cause 10s of millions of dollars in potential losses



Solving the Problem



Choosing Features and Labels

Features

Product

Sub Product

Issue

Sub Issue

Consumer Complaint

Company

Submitted Via

Sentiment Score (Engineered)

Prediction label

Timely Response

True or False

Pre-Processing Data



Drop Columns with large amounts of missing data



Remove irrelevant data columns



Fill in missing values

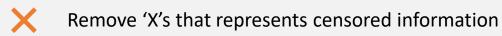


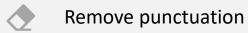
One hot encode categorical features



Apply NLP techniques on written complaints

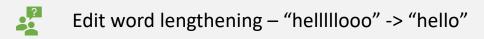
Natural Language Processing







Remove stop words that give little meaning – 'of', 'is'



Tokenize text – create a list of words

Stemming – removing suffixes like "-ly" and "-ing"

Lemmatize – get the root word, "ran", "runs", "running" -> "run"

Feature
Engineering
– Sentiment
Score

Scored written complaints based on the 'attitude'

Used sentiment analysis package called VADER from NLTK

Score between -1 and +1

More negative score means a negative attitude and vice versa

Machine
Learning —
Binary
Classification

Logistic Regression Random Forest





Accuracy is a bad measure ~97% of labels were True for timely responses



We want the model to accurately classify complaints that were not responded to appropriately - < 3% of the data



Instead use log loss, precision, recall, AUC

Why Log Loss?

- If the model classifies label as True for all data, then still 97% accurate
- Log loss measure the error, want lowest score possible
- Log loss punishes incorrect predictions
- Ex) predicts True with 90% probability, but actually False log loss 2.30
- Ex) predicts True with 50% probability, but False log loss 0.69

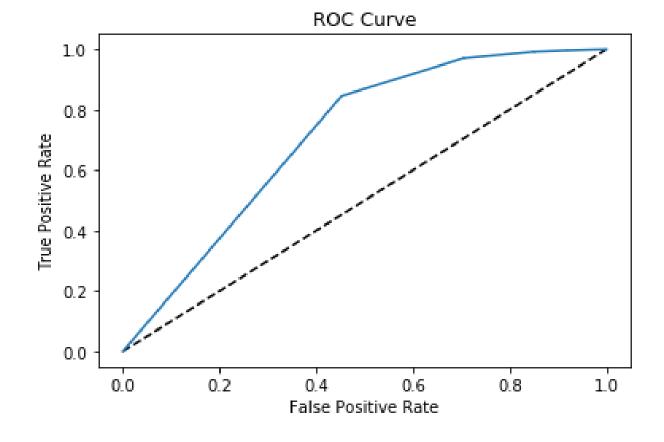


Results of Base Model – Random Forest

Random Forest (Un-tuned)								
	Precision		Recall	call F1-Score		Support		
FALSE		0.44	0.05		0.08	2843		
TRUE		0.97	1		0.98	93117		

Log Loss	Accuracy	AUC
0.5128	0.9699	0.7124

Currently the model predicts a False Label is False 44% of the time, but only correctly identifies 5% of all False labels.

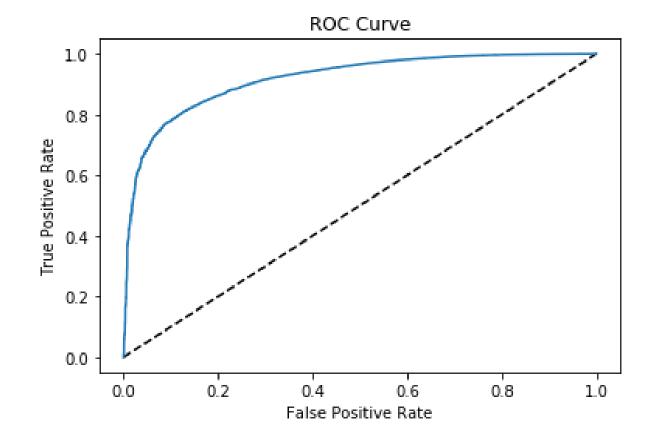


Results of Best Model – Logistic Regression

Logistic Regression (Tuned Hyper Parameters, Lemmatization)								
	Precision	Recall	F1-Score	Support				
FALSE	0.61	0.21	0.31	2843				
TRUE	0.98	1	0.99	93117				

Log Loss	Accuracy	AUC			
0.0875	0.9731	0.9175			

Currently the model predicts a False Label is False 61% of the time, and correctly identifies 21% of all False labels.



Notes on Model Performance

Trained different versions of models with different NLP pre-processing such as with and without lemmatization

Lemmatization showed better results than stemming

Used Bag of Words model – analyzes based on token counts

Adding sentiment score did not improve results

Used grid search and randomized grid search to tune hyper parameters

Satisfied with results?

Better predictions are always better

Being able to identify 21% of hard to handle responses is still a win

Financial products are large dollar items

Failure to rectify some credit or debt complaints can ruin consumer lives

21% of difficult complaints from ~1,400,000 is still 7350 complaints identified

Some Math

- According to https://www.magnifymoney.com/blog/mortgage/u-s-mortgage-market-statistics-2018/, the average mortgage balance is \$148,060
- According to CFPB data in this project: 5528 complaints that were not responded in an appropriate manner were mortgages
- Being able to identify 21% is 1160 complaints
- That's a lot of money and risk to lose due to preventable customer churn

For a large bank that is: \$148,060 X 1160 = \$171,749,600!!!

Business Use Cases

The exploratory data analysis itself has identified that many of these issues stem from mishandling of customer information. A company who sees similar patterns needs to rethink how employees are handling sensitive information. For example, bad credit reporting from mishandling information can cause a consumer to not be approved for a loan, that means a loss for the financial firm as well.

The classification model can be implemented on every written complaint sent to a firm. Any predictions of 'difficult' complaints for a certain product can be handed to a special team to prevent customer churn or better handle a difficult situation. These complaints should not be handled by less experienced employees. Special protocols and procedures can be put in place to handle any complaints tagged as 'difficult' by the model.

Looking Forward

Since NLP and sentiment analysis are not exact sciences, it would be a great exercise to see what improvements can be done to further preprocess the written complaints text

Finding better features to engineer for better performance

Trying out a larger variety of machine learning models

Thank You!

By: Justin Tsao

Find the full project and all the code here:

https://github.com/jltsao88/Improving Customer Service Machine Learning

Examples of Correct Predictions

	product	sub_product	issue	sub_issue	consumer_complaint_narrative	company	submitted_via	True_Label	Prediction
48212	Debt collection	Payday loan	Improper contact or sharing of info	Contacted me after I asked not to	Integrated Services contacted me at my home an	Integrated Recovery Services	Web	False	False
47667	Debt collection	I do not know	Improper contact or sharing of info	Contacted me after I asked not to	calls to phone number reassigned to us 6 month	High Point Asset Inc	Web	False	False
377016	Debt collection	I do not know	Improper contact or sharing of info	Contacted me after I asked not to	I have been contacted several times by an auto	Direct Recovery Services, LLC	Web	False	False
290194	Debt collection	I do not know	Cont'd attempts collect debt not owed	Debt is not mine	I recently pulled my credit and discovered tha	Credit Collections U.S.A., L.L.C.	Web	False	False
21655	Debt collection	Payday loan	Communication tactics	Threatened to take legal action	Continued to call even though I asked several	NPS GROUP	Web	False	False

correct_predict['consumer_complaint_narrative'][48212]

'Integrated Services contacted me at my home and on my personal cell phone about a debt owed by a family member After telling them to longer call me at these numbers and that I have no business with them they continue to call They are calling from'

correct_predict['consumer_complaint_narrative'][47667]

'calls to phone number reassigned to us 6 months ago Previous phone number owner named in complaint Told agent he had wrong number and he agreed to remove us from further calls but calls continue Fed up with their calling'

Examples of Incorrect Predictions

	product	sub_product	issue	sub_issue	consumer_complaint_narrative	company	submitted_via	True_Label	Prediction
299834	Bank account or service	Other bank product/service	Account opening, closing, or management	None	I have owned a dividend reinvestment account f	WELLS FARGO & COMPANY	Web	False	True
30002	Debt collection	Medical	Disclosure verification of debt	Right to dispute notice not received	I have made numerous request for Central Finan	TENET HEALTHCARE CORPORATION	Web	False	True
196760	Credit reporting, credit repair services, or o	Credit reporting	Incorrect information on your report	Account information incorrect	United acceptance is reporting a charge off au	UNITED ACCEPTANCE INC	Web	False	True
357915	Debt collection	I do not know	Disclosure verification of debt	Right to dispute notice not received	i received an update on my credit report and I	Midwest Recovery Systems	Web	False	True
143059	Debt collection	Other debt	Threatened to contact someone or share informa	Talked to a third-party about your debt	On 2018 I received a phone call from East Poin	Eastpoint Recovery Group, Inc	Web	False	True

incorrect_predict['consumer_complaint_narrative'][299834]

'I have owned a dividend reinvestment account for several years This year the administering agency sold the account to Wells Fa rgo Bank When I tried to retrieve my 2016 for this account from Wells Fargo I was presented with an online agreement to waive my rights to receive hard copy tax documents in order to access my I have been advised that this agreement violates existing fed eral law and I consequently refused to sign the online waiver Your assistance is requested to force Wells Fargo to provide me we ith hard copies of my annual tax documents free of charge'

incorrect_predict['consumer_complaint_narrative'][30002]

'I have made numerous request for Central Financial Control Syndicated surrounding to accounts the creditor has placed on my creditor reports prior to notifying me of the alleged debt as well as refusing to disclose verification of debt The latest communication was sent certified on 2015 and the creditor has refused to respond as well as provide any disclosure of verification of debt details along with supporting documents'

Credit for images

- Slide 2 https://due.com/blog/this-is-the-sad-truth-about-how-most-banks-make-money-and-what-you-can-do-about-it/
- Slide 3 https://www.salesforce.com/ca/blog/2017/02/improve-customer-service.html
- Slide 4 https://www.ratchetandwrench.com/articles/5045-dealing-with-customer-complaints
- Slide 6 https://dribbble.com/shots/2379584-Consumers