
Sub-Product Category Classification

Young Sim, Ruowang Li, Yuchen Xu, Xian Qu



Table of Contents

Overview

Project objective

Understanding the Clients

Challenges

Deliverables

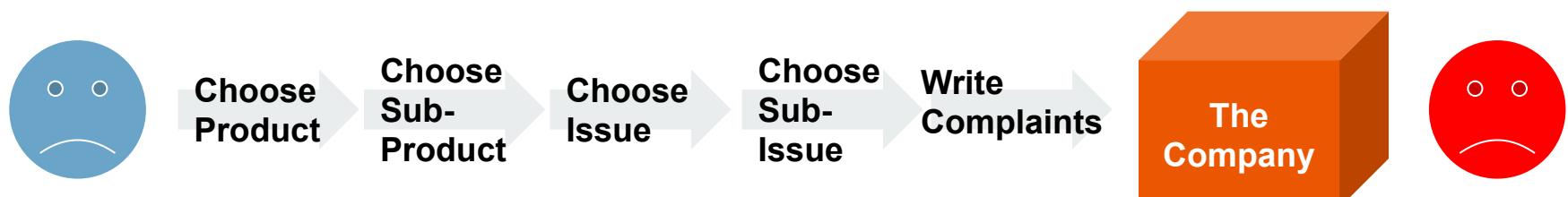
Visualizations

Proposed solution

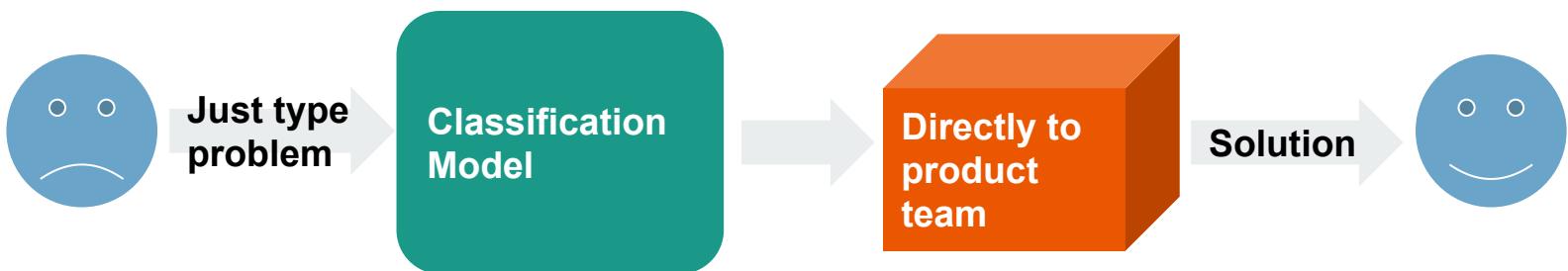
Development



Overview: Before



Overview: What we did





Understanding the Clients

Who are the people involved?

Identify who is involved in this complaint.

- "Just you" if you are the account holder.
- "You and someone else" if you are want to include another account holder.
- "Someone else" if you are submitting on behalf of an authorized third party, such as a lawyer or attorney.

Just you

Someone else

Whic

Un

Ac

Your contact information

Enter information about the account holder or borrower.

First name

Middle (optional)

Last

Yes

Country

City

State

ZIP or postal code

Select an option

Age (optional)

Phone number (optional)

Email

We'll use this email address to send updates about the status of this complaint. It will also be the username for the account.

You don't have an email address.

Have
com

What affiliations does the primary consumer have? Choose all that apply.

(optional)

We use this information to help identify trends in the marketplace.

A servicemember or veteran

A spouse or dependent of a servicemember or veteran

any and ask for a response.

company find you in their system and

Billing address

ZIP code

...

naliman/l



Project objective

1. Triage complaint narratives to departments in charge (classification)
2. Save clients from tedious trouble-shooting forms

Challenges we met

1 Predicting the issues?

Duplicate issue labels under different product labels.

Sub-product is under one Product category

2 Update and revision of categories.

Differently named labels in different times as category names were updated.

3 Business value = Accuracy

If we want this model to improve efficiency in business operation, we need accuracy of automatic classification that beats human choice.

4 Large dataset: expensive to run

For tf-idf feature extraction, cannot consider n-grams except unigrams.

Cannot employ the whole dataset for the 4 models

Limit the number of epochs for the LSTM



Proposed solution

- Predict sub-product types using narratives only
- Build different machine learning models and choose the one with the best accuracy

Process



0 1 Data exploration and pre-processing

Use the latest categories that were last updated in 2017.

Feature extraction

Use word vectors and tf-idf



0 2



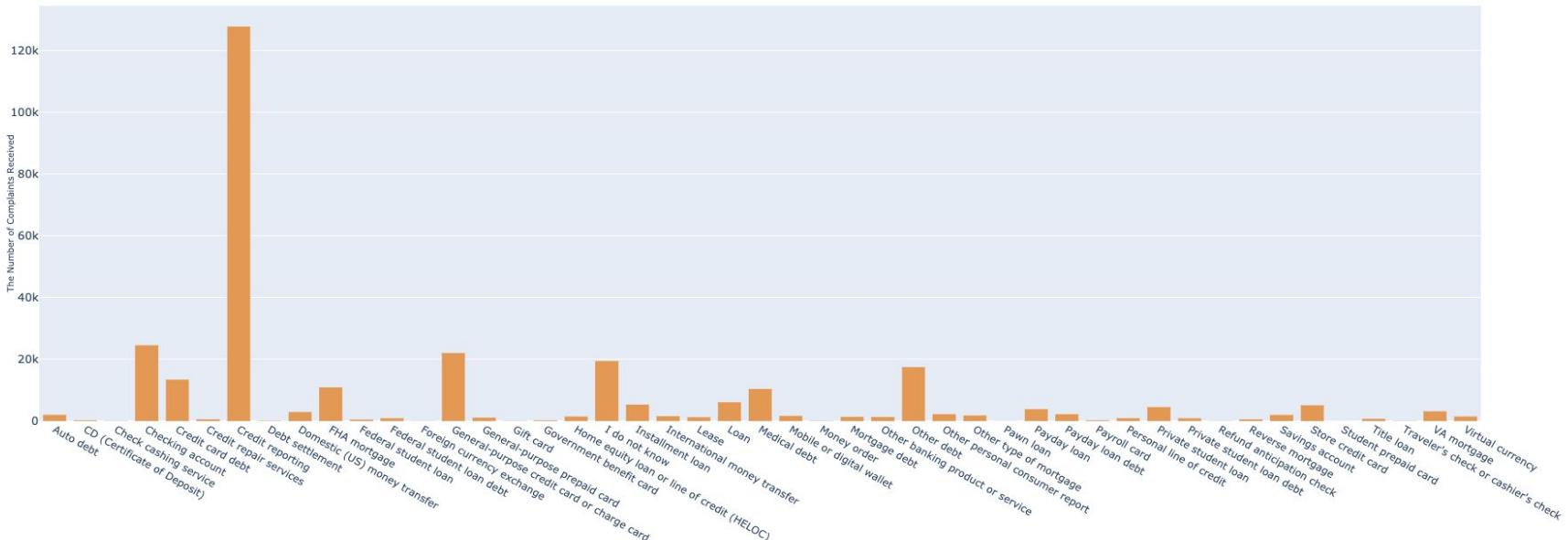
0 3 Models and Evaluation

- LSTM
- Logistic Regression
- SVM
- Naive Bayes
- Random Forest

The number of complaints received per sub-product is imbalanced:
accuracy and F1 scores for evaluation

Distribution

The Number of Complaints Received by Sub-Product





LSTM Model (3 Epochs)

Accuracy

74%

Pretrained word vectors (fasttext provided by facebook) were implemented. Words not included in the pretrained vectors were set to weight 0. More than 50% of the words from our data were not included in the pretrained data.

F1 Weighted
Avg

72%

The model achieved the performance with less than 50% of the data. So there is a big room to grow with some modifications, such as using other pretrained data that works better for our purpose.

	precision	recall	f1-score	support
Other debt	0.52	0.51	0.52	609
Credit reporting	0.83	0.96	0.89	5854
Checking account	0.67	0.83	0.74	585
General-purpose credit card or charge card	0.68	0.78	0.73	859
International money transfer	0.77	0.29	0.42	35
Other banking product or service	0.00	0.00	0.00	53
I do not know	0.57	0.42	0.48	410
Loan	0.62	0.49	0.55	198
Other type of mortgage	0.27	0.08	0.13	48
Payday loan	0.58	0.28	0.38	50
Private student loan debt	1.00	0.03	0.06	30
Mortgage debt	0.83	0.22	0.34	46
FHA mortgage	0.58	0.73	0.65	161
Credit card debt	0.54	0.50	0.52	491
Mobile or digital wallet	0.45	0.37	0.40	79
Store credit card	0.62	0.36	0.46	203
Savings account	0.75	0.14	0.24	43
Medical debt	0.71	0.52	0.60	366
VA mortgage	0.70	0.25	0.37	63
Installment loan	0.31	0.12	0.18	82
Money order	0.00	0.00	0.00	4
Home equity loan or line of credit (HELOC)	0.71	0.25	0.37	40
Other personal consumer report	0.69	0.10	0.18	89
Auto debt	1.00	0.04	0.08	74
Private student loan	0.72	0.47	0.57	87
Federal student loan debt	0.57	0.16	0.25	25
Personal line of credit	0.00	0.00	0.00	37
Payday loan debt	0.67	0.13	0.22	77
General-purpose prepaid card	0.89	0.44	0.59	39
Credit repair services	0.00	0.00	0.00	26
Traveler's check or cashier's check	0.00	0.00	0.00	3
Domestic (US) money transfer	0.30	0.15	0.20	81
Virtual currency	0.30	0.14	0.19	21
CD (Certificate of Deposit)	0.80	0.24	0.36	17
Title loan	0.00	0.00	0.00	19
Refund anticipation check	0.00	0.00	0.00	2
Debt settlement	0.00	0.00	0.00	4
Check cashing service	0.00	0.00	0.00	6
Lease	0.87	0.30	0.45	43
Government benefit card	1.00	0.05	0.09	22
Reverse mortgage	0.00	0.00	0.00	12
Payroll card	0.00	0.00	0.00	5
Gift card	0.50	0.50	0.50	2
Foreign currency exchange	0.00	0.00	0.00	5
accuracy			0.75	11005
macro avg	0.48	0.25	0.29	11005
weighted avg	0.72	0.75	0.72	11005

Logistic

Accuracy

Cross Validation

75% 74%

F1 Weighted Avg

72%

Linear SVM

Accuracy Cross Validation

73% 72%

F1 Weighted Avg

72%

Linear Support Vector Machines Classification Report

	precision	recall	f1-score	support
Other debt	0.50	0.51	0.51	609
Credit reporting	0.87	0.94	0.90	5854
Checking account	0.70	0.75	0.72	585
General-purpose credit card or charge card	0.65	0.69	0.67	859
International money transfer	0.56	0.40	0.47	35
Other banking product or service	0.17	0.08	0.10	53
I do not know	0.50	0.44	0.47	410
Loan	0.51	0.46	0.49	198
Other type of mortgage	0.21	0.17	0.19	48
Payday loan	0.29	0.22	0.25	50
Private student loan debt	0.29	0.07	0.11	30
Mortgage debt	0.50	0.26	0.34	46
FHA mortgage	0.59	0.65	0.62	161
Credit card debt	0.48	0.50	0.49	491
Mobile or digital wallet	0.39	0.35	0.37	79
Store credit card	0.49	0.43	0.46	203
Savings account	0.40	0.23	0.29	43
Medical debt	0.59	0.54	0.57	366
VA mortgage	0.51	0.35	0.42	63
Installment loan	0.22	0.12	0.16	82
Money order	0.33	0.25	0.29	4
Home equity loan or line of credit (HELOC)	0.40	0.30	0.34	40
Other personal consumer report	0.40	0.21	0.28	89
Auto debt	0.29	0.09	0.14	74
Private student loan	0.67	0.48	0.56	87
Federal student loan debt	0.32	0.24	0.27	25
Personal line of credit	0.25	0.05	0.09	37
Payday loan debt	0.43	0.23	0.30	77
General-purpose prepaid card	0.59	0.41	0.48	39
Credit repair services	0.22	0.08	0.11	26
Traveler's check or cashier's check	0.00	0.00	0.00	3
Domestic (US) money transfer	0.27	0.15	0.19	81
Virtual currency	0.25	0.24	0.24	21
CD (Certificate of Deposit)	0.82	0.53	0.64	17
Title loan	0.40	0.11	0.17	19
Refund anticipation check	0.00	0.00	0.00	2
Debt settlement	0.00	0.00	0.00	4
Check cashing service	0.00	0.00	0.00	6
Lease	0.55	0.40	0.46	43
Government benefit card	0.83	0.23	0.36	22
Reverse mortgage	0.80	0.33	0.47	12
Payroll card	0.00	0.00	0.00	5
Gift card	0.50	0.50	0.50	2
Foreign currency exchange	0.00	0.00	0.00	5
accuracy			0.73	11005
macro avg	0.40	0.30	0.33	11005
weighted avg	0.71	0.73	0.72	11005

Naive Bayes

Accuracy Cross Validation

68% **66%**

F1 Weighted Avg

63%

	precision	recall	f1-score	support
Other debt	0.46	0.48	0.47	609
Credit reporting	0.82	0.90	0.86	5854
Checking account	0.52	0.79	0.63	585
General-purpose credit card or charge card	0.42	0.80	0.55	859
International money transfer	0.00	0.00	0.00	35
Other banking product or service	0.00	0.00	0.00	53
I do not know	0.64	0.28	0.39	410
Loan	0.49	0.40	0.44	198
Other type of mortgage	0.00	0.00	0.00	48
Payday loan	0.50	0.02	0.04	50
Private student loan debt	0.00	0.00	0.00	30
Mortgage debt	0.67	0.04	0.08	46
FHA mortgage	0.48	0.68	0.56	161
Credit card debt	0.43	0.45	0.44	491
Mobile or digital wallet	0.83	0.06	0.12	79
Store credit card	0.36	0.04	0.08	203
Savings account	0.00	0.00	0.00	43
Medical debt	0.75	0.34	0.46	366
VA mortgage	1.00	0.06	0.12	63
Installment loan	0.00	0.00	0.00	82
Money order	0.00	0.00	0.00	4
Home equity loan or line of credit (HELOC)	0.00	0.00	0.00	40
Other personal consumer report	0.12	0.01	0.02	89
Auto debt	0.00	0.00	0.00	74
Private student loan	0.62	0.34	0.44	87
Federal student loan debt	0.00	0.00	0.00	25
Personal line of credit	0.00	0.00	0.00	37
Payday loan debt	0.00	0.00	0.00	77
General-purpose prepaid card	0.00	0.00	0.00	39
Credit repair services	0.00	0.00	0.00	26
Traveler's check or cashier's check	0.00	0.00	0.00	3
Domestic (US) money transfer	0.00	0.00	0.00	81
Virtual currency	1.00	0.05	0.09	21
CD (Certificate of Deposit)	0.00	0.00	0.00	17
Title loan	0.00	0.00	0.00	19
Refund anticipation check	0.00	0.00	0.00	2
Debt settlement	0.00	0.00	0.00	4
Check cashing service	0.00	0.00	0.00	6
Lease	0.00	0.00	0.00	43
Government benefit card	0.00	0.00	0.00	22
Reverse mortgage	0.00	0.00	0.00	12
Payroll card	0.00	0.00	0.00	5
Gift card	0.00	0.00	0.00	2
Foreign currency exchange	0.00	0.00	0.00	5
accuracy			0.68	11005
macro avg	0.23	0.13	0.13	11005
weighted avg	0.64	0.68	0.63	11005

Random Forest

Accuracy Cross Validation

53% 53%

F1 Weighted Avg

37%

Random Forest Classification Report

	precision	recall	f1-score	support
Other debt	0.00	0.00	0.00	609
Credit reporting	0.53	1.00	0.69	5854
Checking account	0.00	0.00	0.00	585
General-purpose credit card or charge card	0.00	0.00	0.00	859
International money transfer	0.00	0.00	0.00	35
Other banking product or service	0.00	0.00	0.00	53
I do not know	0.00	0.00	0.00	410
Loan	0.00	0.00	0.00	198
Other type of mortgage	0.00	0.00	0.00	48
Payday loan	0.00	0.00	0.00	50
Private student loan debt	0.00	0.00	0.00	30
Mortgage debt	0.00	0.00	0.00	46
FHA mortgage	0.00	0.00	0.00	161
Credit card debt	0.00	0.00	0.00	491
Mobile or digital wallet	0.00	0.00	0.00	79
Store credit card	0.00	0.00	0.00	203
Savings account	0.00	0.00	0.00	43
Medical debt	0.00	0.00	0.00	366
VA mortgage	0.00	0.00	0.00	63
Installment loan	0.00	0.00	0.00	82
Money order	0.00	0.00	0.00	4
Home equity loan or line of credit (HELOC)	0.00	0.00	0.00	40
Other personal consumer report	0.00	0.00	0.00	89
Auto debt	0.00	0.00	0.00	74
Private student loan	0.00	0.00	0.00	87
Federal student loan debt	0.00	0.00	0.00	25
Personal line of credit	0.00	0.00	0.00	37
Payday loan debt	0.00	0.00	0.00	77
General-purpose prepaid card	0.00	0.00	0.00	39
Credit repair services	0.00	0.00	0.00	26
Traveler's check or cashier's check	0.00	0.00	0.00	3
Domestic (US) money transfer	0.00	0.00	0.00	81
Virtual currency	0.00	0.00	0.00	21
CD (Certificate of Deposit)	0.00	0.00	0.00	17
Title loan	0.00	0.00	0.00	19
Refund anticipation check	0.00	0.00	0.00	2
Debt settlement	0.00	0.00	0.00	4
Check cashing service	0.00	0.00	0.00	6
Lease	0.00	0.00	0.00	43
Government benefit card	0.00	0.00	0.00	22
Reverse mortgage	0.00	0.00	0.00	12
Payroll card	0.00	0.00	0.00	5
Gift card	0.00	0.00	0.00	2
Foreign currency exchange	0.00	0.00	0.00	5
accuracy			0.53	11005
macro avg	0.01	0.02	0.02	11005
weighted avg	0.28	0.53	0.37	11005



Future works: Model Improvement

Calibrate the parameters of Logistic Regression and Linear SVM models

Provide more training for LSTM model

Find other pre-trained data that works better for our purpose

Imbalanced data: Oversampling/ SMOTE...



Future works: Business Value

Automatically update the
obsolete database



Issue prediction under
each sub-product
category

Further improvement of
classification accuracy

Category updates



Thank you

