Credit complaints classification

Classification of product associated with credit complaint with the use of Topic Modelling

Jan Frąckowiak

Objective

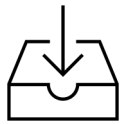
The goal of this project was classification of product associated with credit complaints

Data encompassed 4 distinct products categories:

- Checking or savings account
- Credit card or prepaid card
- Debt collection
- Mortgage

Potential business benefits:

building categorised dataset for product-related investigations
  - identifying risks and space for products' development
  - assessing company's compliance with regulations



operational efficiency (complaints arrive at target)

Dataset

Perfectly balanced data featuring 120 000 Consumer complaints related to four most frequently reported financial products.

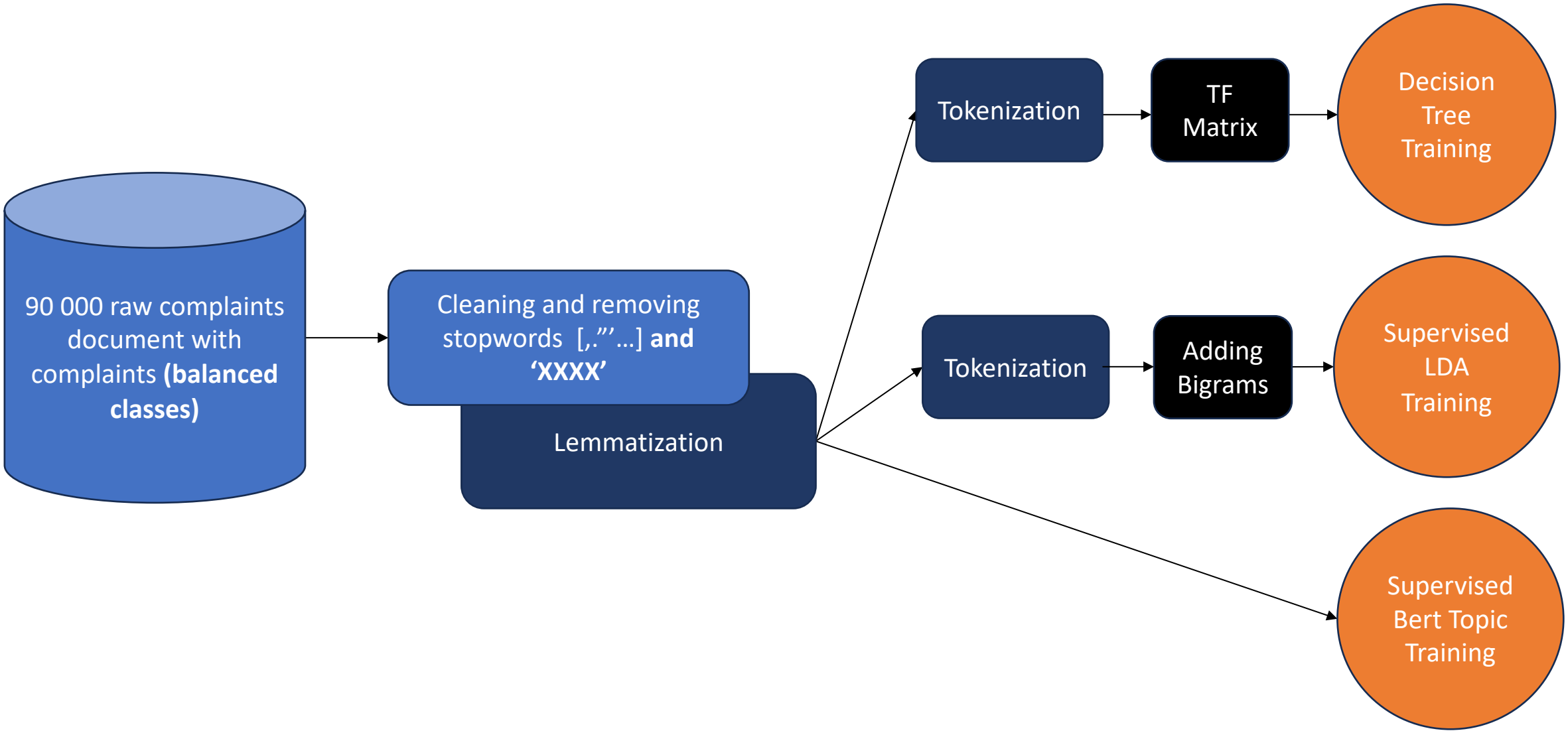
Complaints were mostly sent to companies. by email.

Documents of length:
Min: 3
Max 21381
Median: 524

Source: https://www.consumerfinance.gov/data-research/consumer-complaints/

Checking or savings account	Credit card or prepaid card	Debt collection	Mortgage
On XX/XX/2020, Chime bank placed a hold on my account and ended up closing the account that day citing section 9 of their service agreement.	I currently have an apple card. I was going to make a purchase on XXXX but the seller said that he could have me make the purchase on his website and it would be cheaper for shipping and taxes. (...)	(...) I attempted to negotiate a 5 year payoff with XXXX and they refused after my home was damaged in Hurricane Sandy (...)	In XXXX I refinanced my mortgage loan with TCF Bank. In XXXX TCF sold my loan to XXXX (XXXX XXXX XXXX). When I received loan documents from XXXX a (redacted) trust affiliated with XXXX XXXX was listed as lien holder. (...)

Preparing data for classification



Training results

Model	Hyperparams	Multilabel ROC AUC on test	Balanced Accuracy on test
Decision Tree (Benchmark)	Obtained with 3- fold Grid Search: {'max_depth': 20, 'min_samples_leaf': 4, 'min_samples_split': 10}	87.59%	81.37%
SLDA with response variable	Obtained with 5-fold CV: {'k'=30, 'min_df'=30, 'rm_top'=0, 'alpha'=0.1, 'eta'=0.01, 'mu'=0, 'nu_sq'=1, 'glm_param'=1}	96.9%	87.6%
BERT Topic with Logit	reduce_frequent_words = True	93.3% (calculated on labels not probs)	89.9%

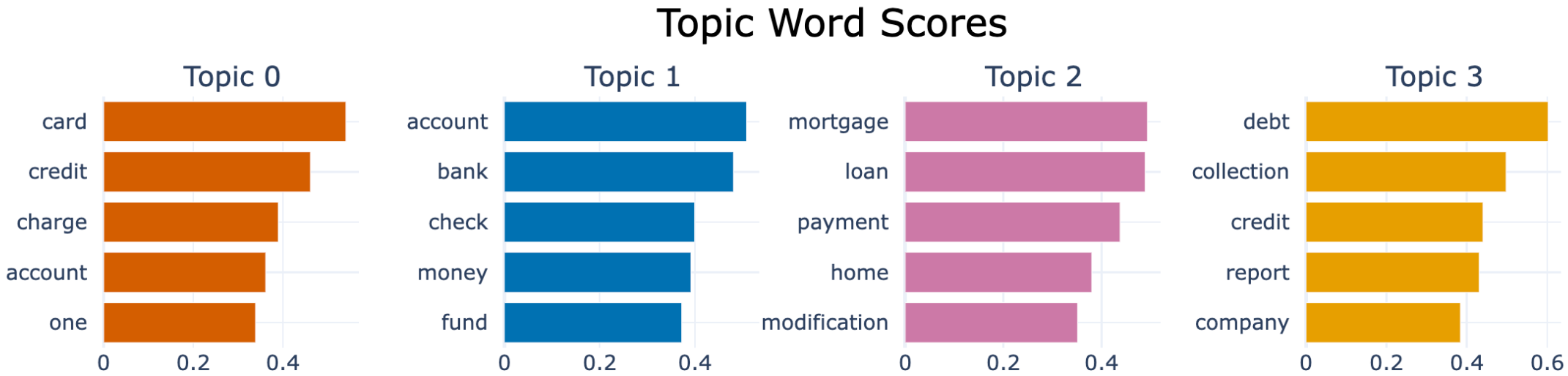
Most important +/- topic coefficients (SLDA)

Checking or savings account	Topic 15 (-23.7)	Topic 29 (-15.9)	Topic 9 (-13.4)	Topic 6* (+12.15)
Credit card or prepaid card	Topic 15 (+23.8)	Topic 9 (-17.9)	Topic 29 (-16.5)	Topic 28 (-14.9)
Debt collection	Topic 7 (+18.1)	Topic 13 (+16.9)	Topic 15 (-12.4)	Topic 29 (-11.7)
Mortgage	Topic 9 (+21.3)	Topic 29 (+19.1)	Topic 28 (+16.2)	Topic 7 (-15.8)

* Topic 6 was the 6th most important topic

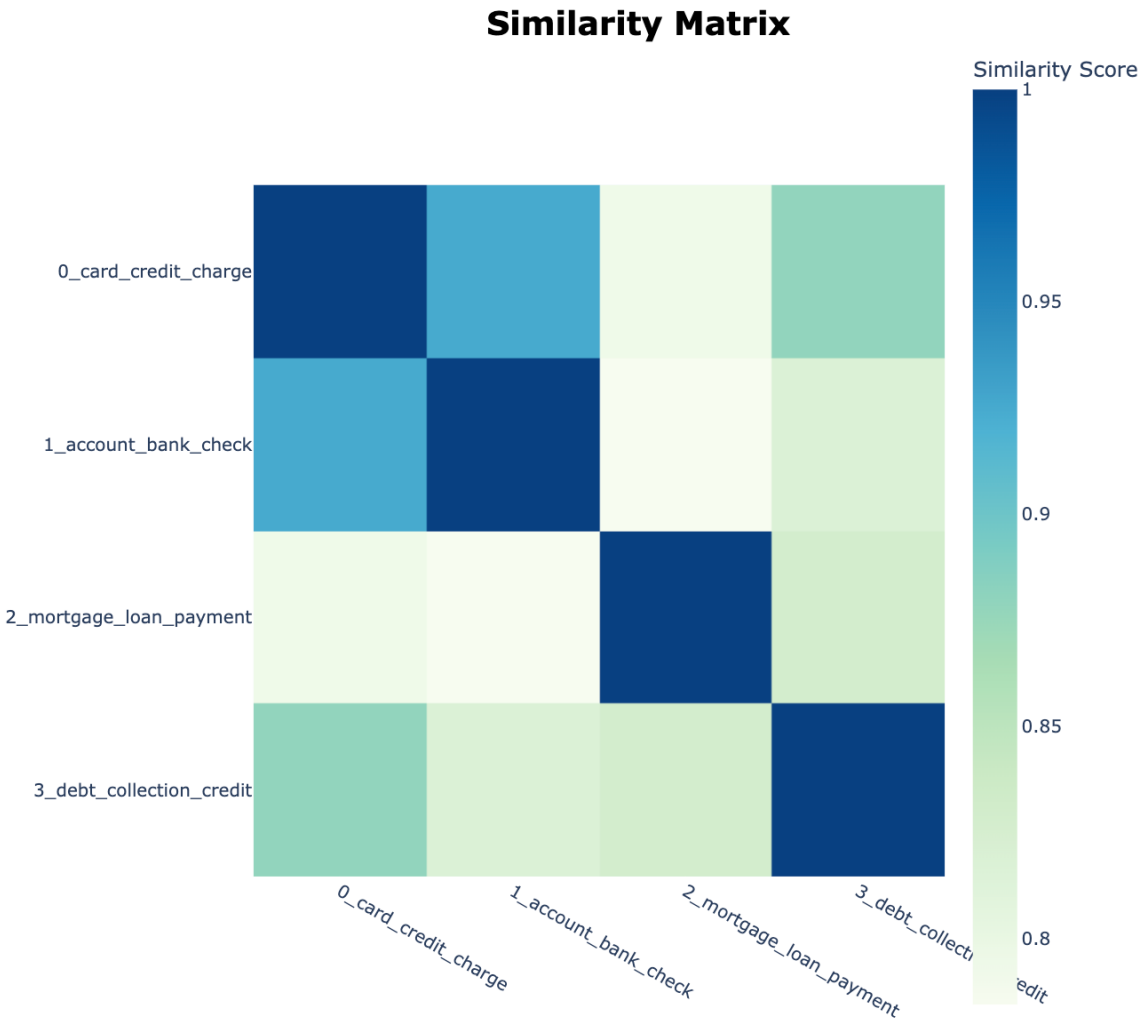
Topic 6	Topic 7	Topic 9	Topic 13	Topic 15	Topic 28	Topic 29
told	debt	mortgage	debt	card	loan	loan
would	collection	insurance	company	credit	mortgage	mortgage
called	credit	company	call	credit card	rate	modification
said	company	mr	calling	score	home	foreclosure
back	report	mort. com.	collect	limit	closing	home

Topic Representations (Bert)



Topic	Documents Count	Bert Representation	Class
0	2410	0_card_credit_charge_account	Credit card or prepaid card
1	24044	1_account_bank_check_money	Checking or savings account
2	24012	2_mortgage_loan_payment_home	Mortgage
3	23837	3_debt_collection_credit_report	Debt collection

Relationship between topics (Bert)



Thank you!