Text Mining on Consumer Complaints in Financial Products

Jiahui Bi, Xinlian Huang, Bowen Lu, Lulu Zhu
Instructor: Rong(Emily) Liu

Business Intelligence & Analytics
November 13th 30th, 2017

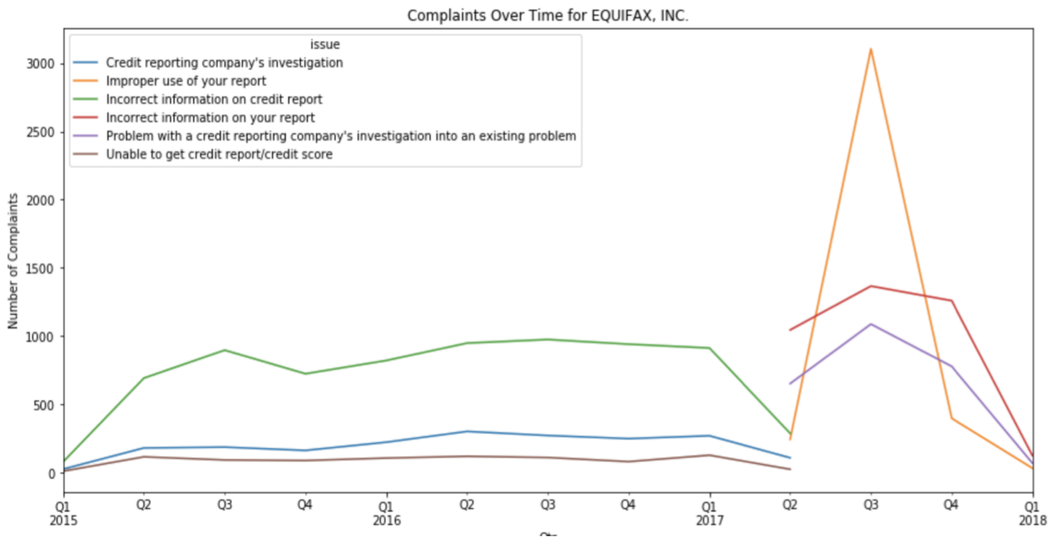
Introduction

- Consumer Financial Protection Bureau: collect customers complaints regarding financial products to promote fairness and transparency for financial products and services.
- Motivation:
  - Simplify complaints submission process and improve customer experience
  - Refine issue classification process

Objective:

- Build text auto-classification system, complaints can be accurately assigned to specific department or team.

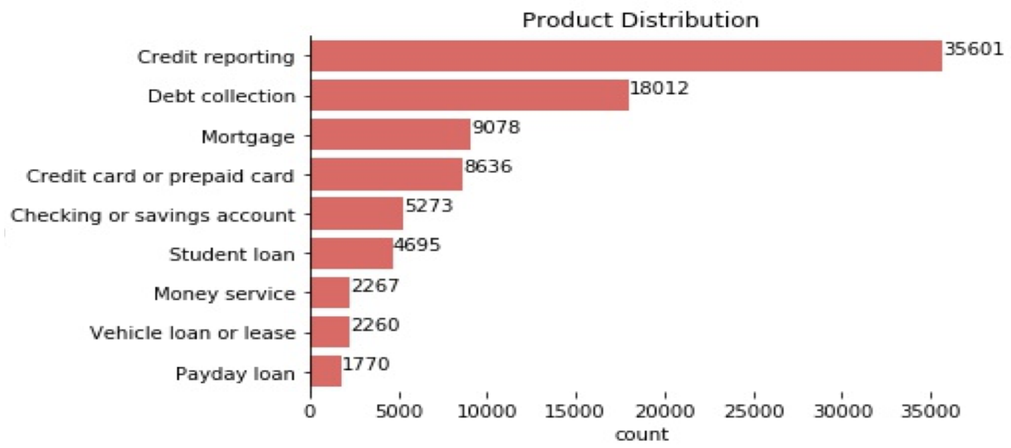
Timeline Diagnosis



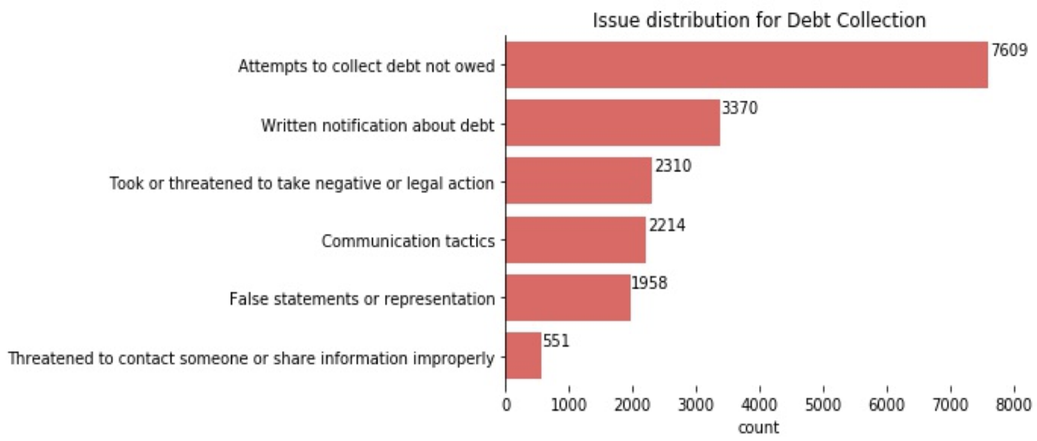
- Equifax data breach scandal in September 2017
- CFPB rebuilt its' issues classification system

Labels Preprocessing

- Products Distribution



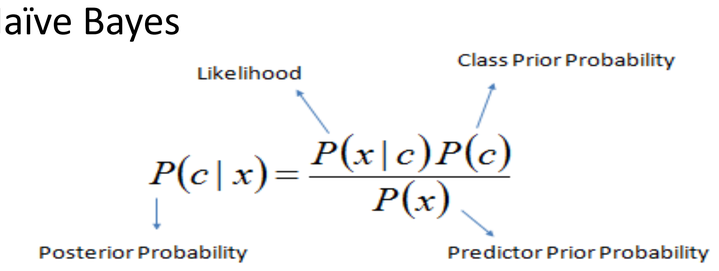
- Issues Distribution



- Relabel
  - Combine "student loans", "vehicle loans" and "payday loans" as "loans"
  - Combine low amount issues with chaotic information as "others"
  - Combine product and issue as new label

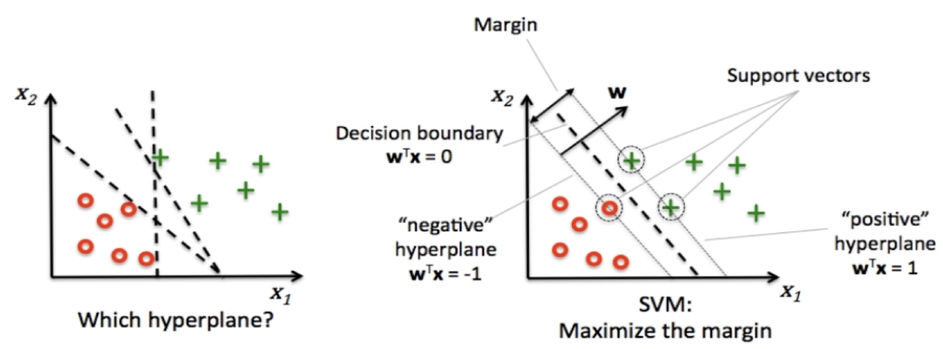
Modeling

- Naïve Bayes

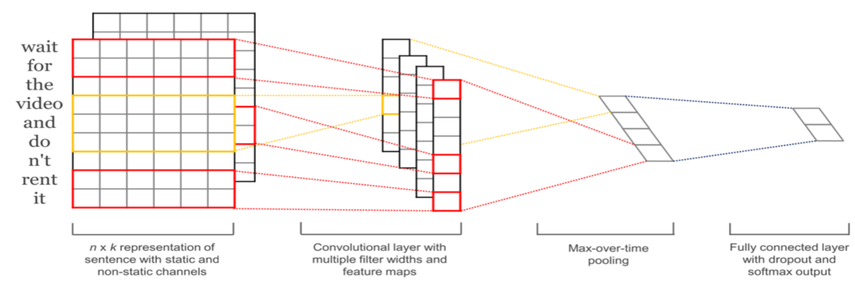


$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- Support Vector Machine



- Convolutional Neural Network

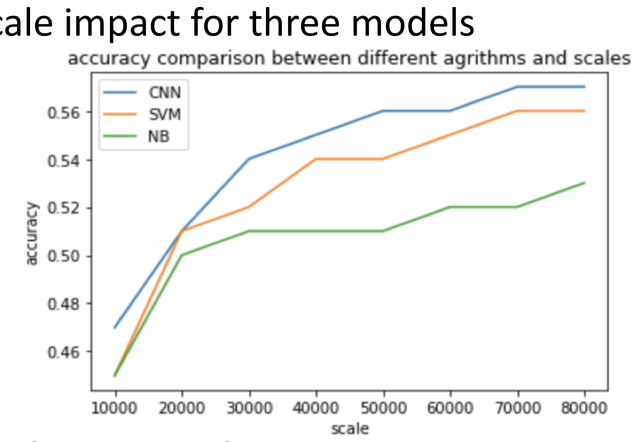


Evaluation

- Embedding methods for CNN

Word Vector Weight	No pre-trained weight	Pre-trained weight + not trainable	Pre-trained weight + trainable
val_accuracy	0.5744	0.5649	0.5780

  - **No pre-trained weight:** train the word vector from all 0.
  - **Pre-trained weight + not trainable:** use pre-trained word vector from Glove, weight is not changeable in the training process.
  - **Pre-trained weight + trainable:** use pre-trained word vector from Glove, but the weight changes during the training process.
- Dataset scale impact for three models



Vulnerability Analysis

- Narratives are written in an oral way
- Labels are overlapped with each other
  - "money transfer" relates to "bank account"

Improvement

- Two-step Classifier: Predict the product type, within the product, predict the issue. Avg. Precision: 0.61

Future Work

- Dimension Reduction: Truncated SVD
- Gradient Boosting Decision Trees