

# bastani\_2019\_latent\_dirichlet\_allocation\_lda\_for\_topic\_modeling\_of\_the\_cfpb\_consumer\_complaints

## Year

2019

## Author(s)

Kaveh Bastani and Hamed Namavari and Jeffrey Shaffer

## Title

Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints

## Venue

Expert Systems with Applications

---

## Topic labeling

Manual

## Focus

Secondary

## Type of contribution

Established approach

## Underlying technique

Manual labeling

## Topic labeling parameters

\

## Label generation

“Automatic labeling of topics is not possible as discovering the topics is an un-supervised learning process. It rather requires human judgments and intervention to examine the coherence and meaningfulness of the topics, and subsequently label them through their judgment.

As a result, after extracting the topics using LDA, the authors validated and labeled the topics, which can be seen in the third column of Table 2. ”

**Table 2**  
Extracted **topics** for the CFPB consumer complaints using LDA.

ID	Topics	Label
0	Payment late made due make appli month past day miss	Payment/late payment
1	Receiv letter sent send mail state request email document notic	Communication
2	Loan student borrow privat navient repay lender default defer forbear	Loan/Student Loan
3	Car vehicl financ dealership dealer ticket book drive trade truck	Auto Loan/Dealership
4	Servic custom repres manag transfer spoke depart supervisor cancel speak	Customer Service
5	Check cash advanc clear return wrote flagstar seiz present payabl	Check
6	File complaint cfpb case complain respond clerk district bsi compliant	CFPB
7	Home hous equiti repair inspect buy door damag sell valu	Home Equity
8	Call phone number person stop time answer messag harass work	Harassment
9	Credit report remov bureau show neg correct inform agenc transunion	Credit Reporting
10	Co program school class colleg signer enrol student region educ	Education
11	Account close open activ statu inaccuraci limit creat past restrict	Account Management
12	Purchas express store buy product order return replac item refund	Store Purchases
13	Consum financi protect practic decept abus institut mislead reposess factor	Consumer Protection
14	Bill insur medic polici hospit cover coverag flood doctor health	Medical Debt/Bill
15	Disput item investig equifax verifi delet inform inaccur valid creditor	Dispute
16	Inquiri author unauthor fraudul tx evict code entitl hard measur	Unauthorized Inquiries
17	Famili mother incom daughter member father live parent employ sister	Family
18	Ident theft fraudul polic victim block ftc stolen affidavit inform	Identity Theft
19	Modif mortgag loan foreclosur home document review incom program trial	MLMF <sup>a</sup>
20	Address name number secur social person live verifi inform licens	Address
21	Violat law act feder fair right note legal assign trust	FDCPA <sup>b</sup>
22	Credit score line limit hard maci pull drop impact affect	Credit Score
23	Agreement close contract apprais sign settlement lender cost agre valu	Settlement
24	Bankruptci file judgment court discharg chapter judgement record public includ	Bankruptcy
25	Interest rate princip refin lower high higher increas principl reduc	Interest Rate
26	Card credit use discov debit reward cancel activ visa declin	Credit and Debit Card
27	Mortgag tax escrow properti insur increas refin sold home ir	Escrow Account
28	Debt collect valid owe collector creditor origin alleg ceas settl	Debt Collector/Issue
29	Attorney court legal employ law firm serv garnish lawyer threaten	Attorney/Legal Actions
30	Husband wife ex divorc heloc name primari child certif spous	Divorce and Ex-spouse
31	Fraud claim investig fraudul alert depart crimin commit activ suspect	Fraud alert
32	Charg fee overdraft refund transact revers merchant waiv assess post	Charged/Overdraft fees
33	Sale properti short lien titl releas buyer sell sold seller	Property Sale
34	Fund money deposit transact branch debit withdraw bank transfer hold	Fund and Deposit
35	Collect agenc recoveri portfolio reinsert invoic enhanc solut central erc	Collection Agency
36	Offer promot point requir bonu term benefit mile condit receiv	Rewards and Promotions
37	System onlin error websit access free site comput log correct	Online Service
38	Paid amount balanc statement owe full show total due payoff	Payoff
39	Applic approv paperwork process appli deni fill delay discrimin denial	Application

<sup>a</sup> Mortgage/Loan Modification and Foreclosure (MLMF).

<sup>b</sup> Fair Debt Collection Practices Act (FDCPA) is a federal law that controls and restricts the activities of third-party debt collectors.

## Motivation

“However, in order to characterize the underlying content of the topics, it would be easier to label them rather than presenting them as combinations of words. ”

---

## Topic modeling

LDA

### Topic modeling parameters

$\alpha$ : 0.1

$\eta$ : 0.1

( $\alpha$  is a topic smoothing parameter, and  $\eta$  is a word/term smoothing parameter)

Nr of topics (K): 40

### Nr. of topics

40

---

## Label

Single or multi word labels manually assigned by the authors

### Label selection

\

### Label quality evaluation

\

### Assessors

\

---

## Domain

Paper: Consumer complaints

Dataset: Consumer complaints

## **Problem statement**

The Consumer Financial Protection Bureau (CFPB), receives and processes consumer complaints pertaining to various financial services.

Every complaint narrative provides insight into problems that consumers are experiencing. Manual review of these documents by human experts is not feasible. This requires an intelligent system to analyze narratives automatically and provide insightful knowledge to the experts.

In this paper, we propose an intelligent approach based on latent Dirichlet allocation (LDA) to analyze the CFPB consumer complaints.

The proposed approach aims to extract latent topics in the CFPB complaint narratives, and explores their associated trends over time.

The time trends will then be used to evaluate the effectiveness of the CFPB regulations and expectations on financial institutions in creating a consumer oriented culture.

## **Corpus**

Origin: Consumer Financial Protection Bureau (CFPB)

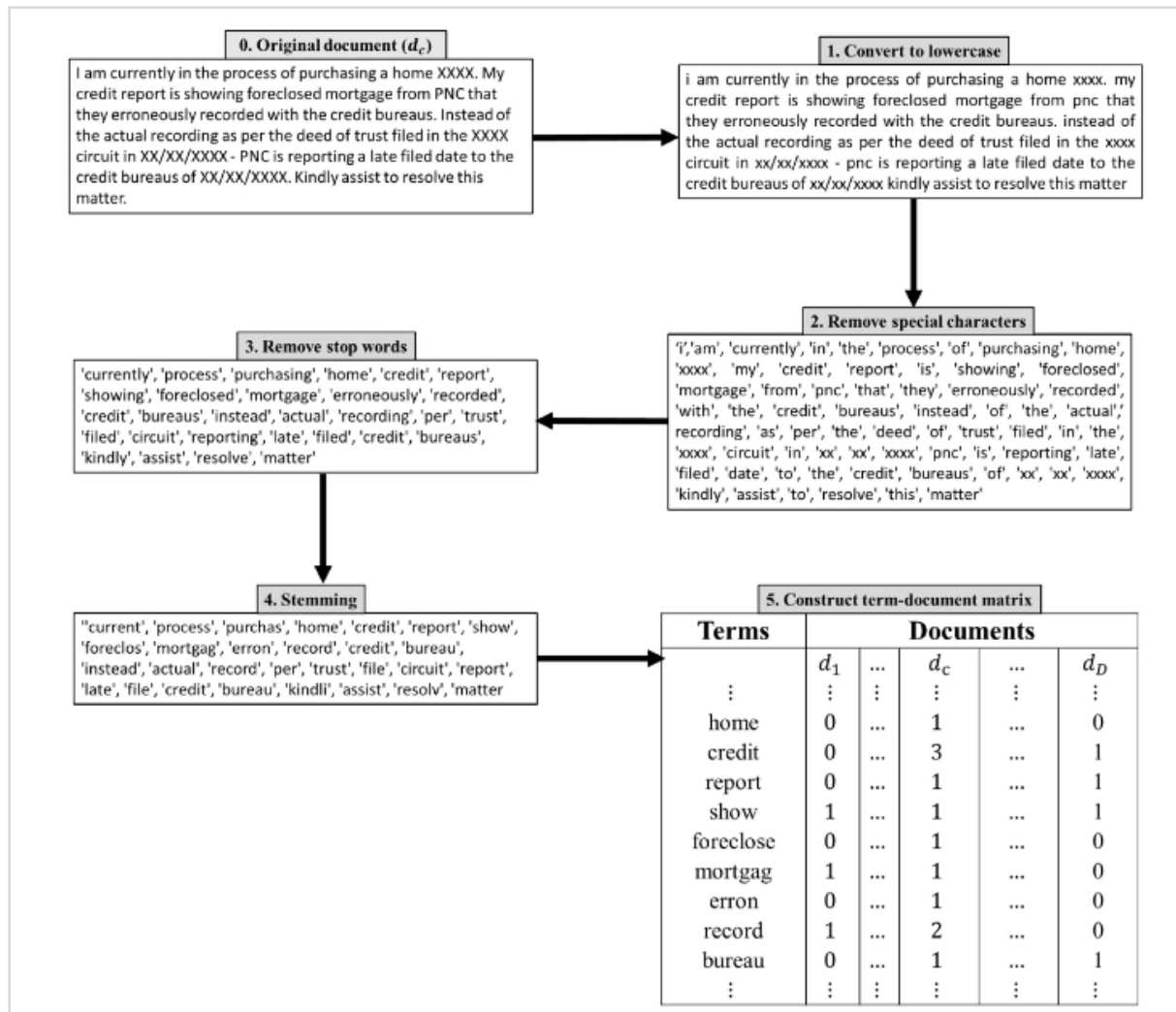
Nr. of documents: 86,803

Details:

## **Document**

Complaint record including certain information such as the description of what happened (the narrative), the date of submission, the product and the issue the complaint is about, the company the complaint was sent to for response

## **Pre-processing**



@article{bastani\_2019\_latent\_dirichlet\_allocation\_lda\_for\_topic\_modeling\_of\_the\_cfpb\_consumer\_complaints,

abstract = {The Consumer Financial Protection Bureau (CFPB), created by congress in 2011, receives and processes consumer complaints pertaining to various financial services. Every complaint narrative provides insight into problems that consumers are experiencing. With increasing number of the CFPB complaint narratives, manual review of these documents by human experts is not feasible. This requires an intelligent system to analyze narratives automatically and provide insightful knowledge to the experts. In this paper, we propose an intelligent approach based on latent Dirichlet allocation (LDA) to analyze the CFPB consumer complaints. The proposed approach aims to extract latent topics in the CFPB complaint narratives, and explores their associated trends over time. The time trends will then be used to evaluate the effectiveness of the CFPB regulations and expectations on financial

institutions in creating a consumer oriented culture. The technology-human partnership between the proposed approach and the CFPB experts could certainly improve consumer experience by providing more efficient and effective investigations of consumer complaint narratives.},

author = {Kaveh Bastani and Hamed Namavari and Jeffrey Shaffer},

date-added = {2023-03-22 17:28:43 +0100},

date-modified = {2023-03-22 17:28:43 +0100},

doi = {https://doi.org/10.1016/j.eswa.2019.03.001},

issn = {0957-4174},

journal = {Expert Systems with Applications},

keywords = {Analytics, Latent Dirichlet allocation, Topic modeling, CFPB, Decision support system, Consumer complaint narratives},

pages = {256-271},

title = {Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints},

url = {https://www.sciencedirect.com/science/article/pii/S095741741930154X},

volume = {127},

year = {2019}}

#Thesis/Papers/Initial