



Park Sangbeom

21800272



Cho Eunsung

21800697



Park Yeeun

22000282

Financial Complaints

Text-Mining Analysis

2023-2 Text Mining Team 1

CONTENTS

- ❑ Introduction
- ❑ EDA
- ❑ Data preprocessing
- ❑ Sentiment analysis
- ❑ Word frequency
- ❑ Classification
- ❑ Conclusion

Problem Definition



Banks record 117% increase in customers' complaints

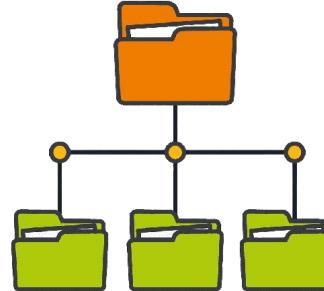
13th October 2023

- Financial companies face a growing volume of customer complaints.
- Prompt and accurate resolution of these complaints is essential.
- However, the increasing number of complaints poses a challenge due to limited company resources.

Project Goal

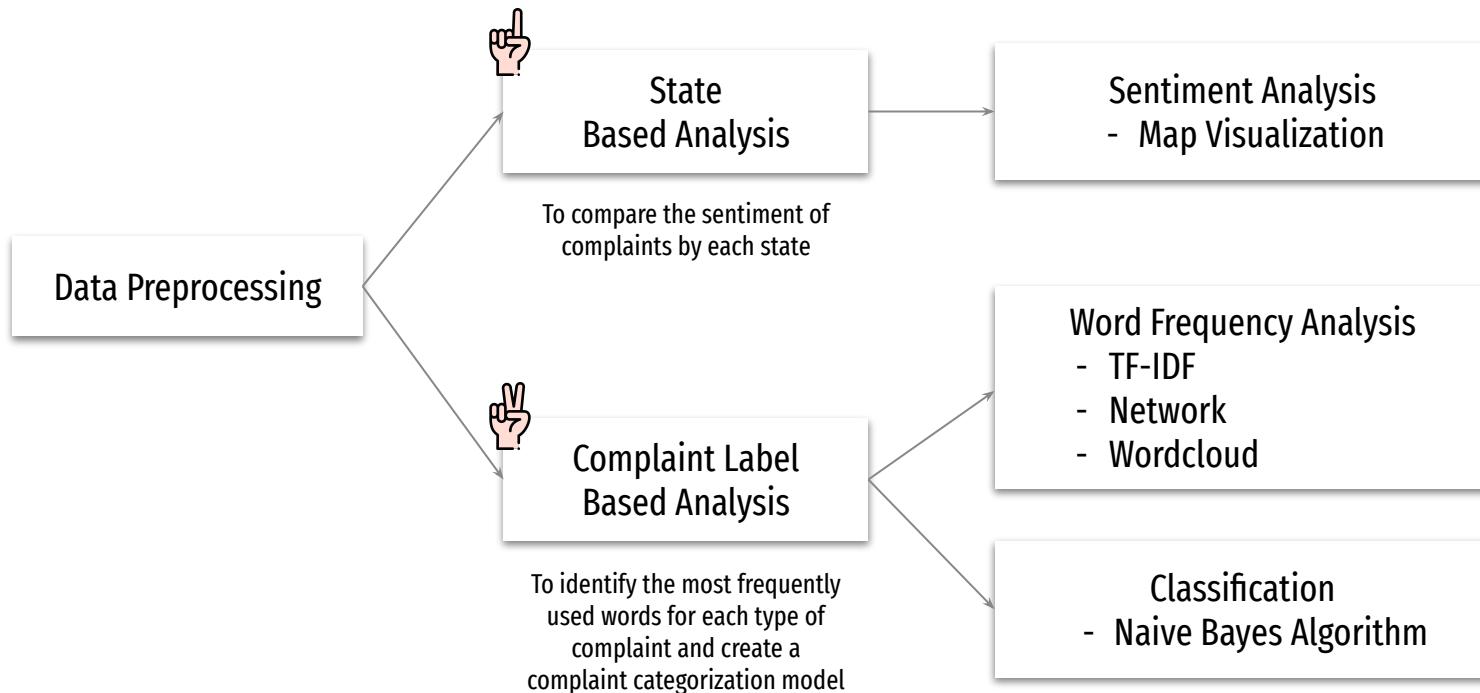


Identify states with the highest complaints and negative sentiments.
Understand specific consumer complaints in those states and focus on them for more efficient resolution.



Analyze word frequency by complaint label, create a machine learning model for efficient classification, enabling effective and fast resolution of increasing complaints.

Analysis Overview



Data Source



Consumer Financial
Protection Bureau

Filter results by...

Date CFPB received the complaint [Show +](#)

Product / sub-product [Show +](#)

Issue / sub-issue [Show +](#)

State [Show +](#)

ZIP code [Show +](#)

Company name [Show +](#)

Did company provide a timely
response? [Show +](#)

Filter



Most popular Bank

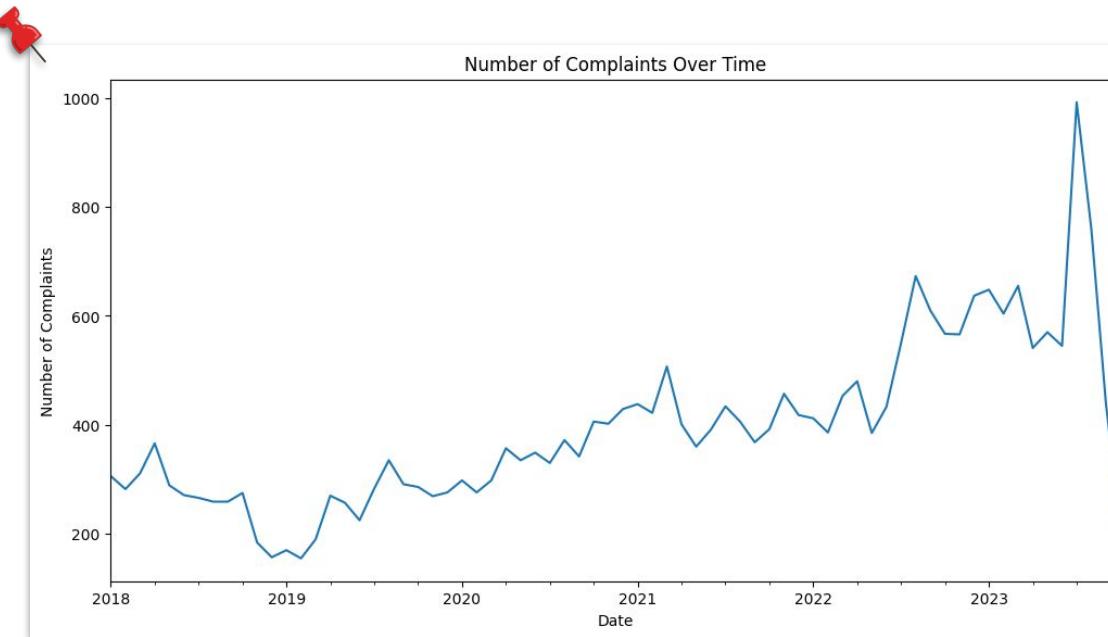


2018~2023

Extracted DataFrame

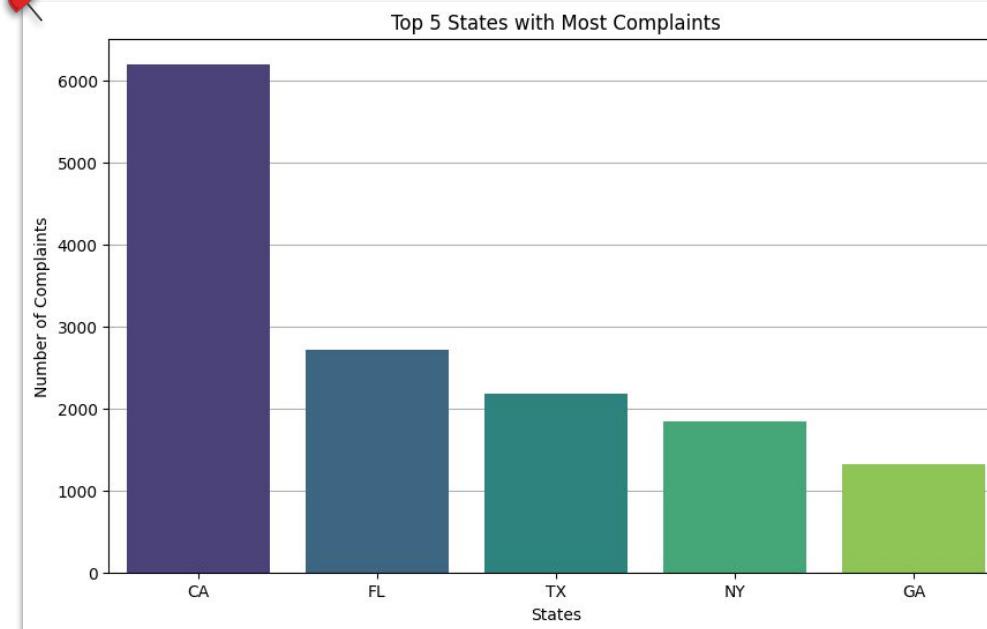
Column-1	Column-2	...	Column-n
		...	
		...	
...
		...	
		...	

Complaints Over Years



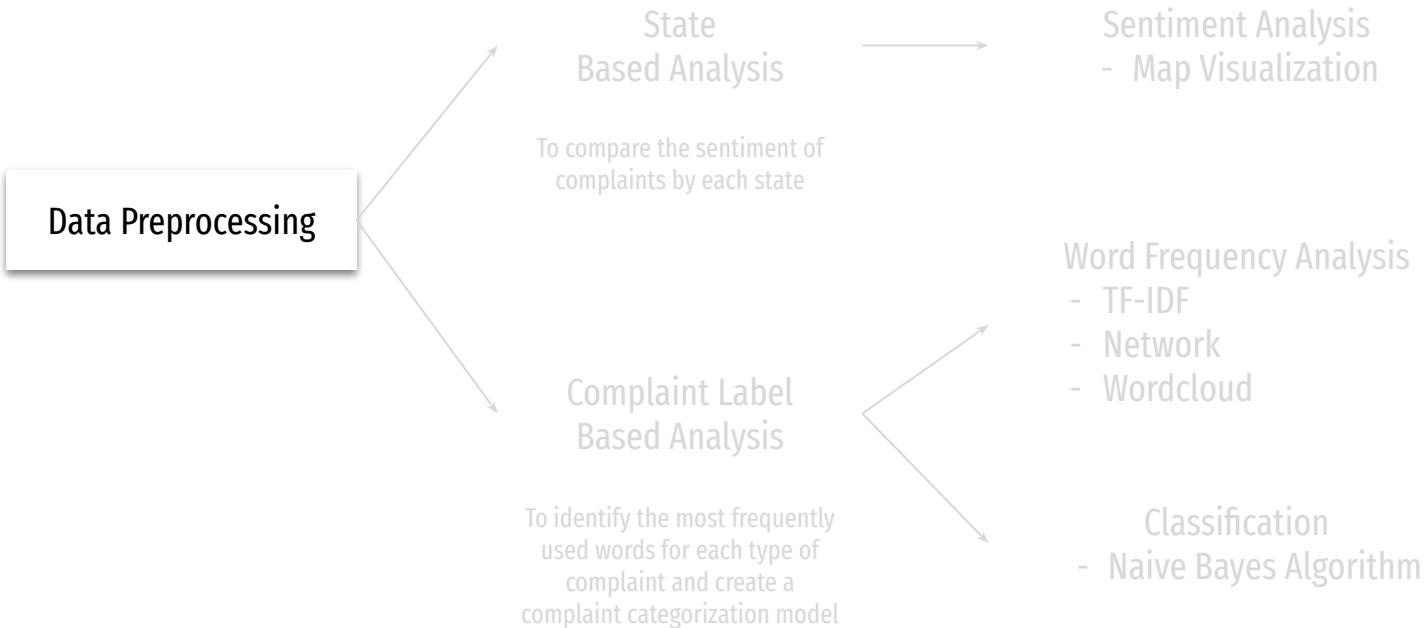
Increase in the number of complaints by year

Top States with Most Complaints



The number of complaints is highest in California, followed by Florida, and Texas.

Analysis Overview

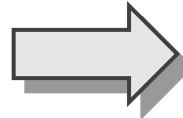


Column Selection

<Original Data>

Date.received
Product
Sub.product
Issue
Sub.issue
Consumer.complaint.narrative
Company.public.response
Company
State
ZIP.code
Tags
Consumer.consent.provided.
Submitted.via
Date.sent.to.company
Company.response.to.consumer
Timely.response.
Consumer.disputed.
Complaint.ID

(27536 * 18)



<Selected Columns>

Date
Label

- Card&Account
- Credit_Management
- Loan

Complaint
State

(27536 * 4)

Column Selection

- Credit card or prepaid card
- Credit card
- Prepaid card
- Checking or savings account
- Money transfer, virtual currency, or money service

Card&Account

- Credit reporting, credit repair services, or other personal consumer reports
- Credit reporting or other personal consumer reports
- Debt or credit management
- Debt collection

Credit Management

- Mortgage
- Vehicle loan or lease
- Payday loan, title loan, or personal loan
- Student loan
- Payday loan, title loan, personal loan, or advance loan

Loan

Preprocessing Steps



Corpus

Text Pre-processing

DTM

text

1 Dear Agent of the CFPB, or to whom it may concern, This message regards obstruction by Bank of America, which is attempting to account that I am under no obligation to maintain. \n\nIn XXXX, I contacted my mortgage company, Bank of America, to request that the mortgage be cancelled. They refused, because I had just switched home-owner's insurance and they needed proof of insurance, which was placed policy. At that time, which was Monday, XXXX XXXX, I was in contact with XXXX XXXX (XXXX), and she told me that I would need to provide proof of payment to prevent the lender-placed policy from being executed. Without a lender-placed policy, I would be free to agree to that Friday (XXXX XXXX) as a deadline, because that was the date they had scheduled to send the letter to give notice. I would provide the required documentation and she would see to it that the escrow account was closed. \n\nMeanwhile, directly, which was done by his employee XXXX XXXX (XXXX), who requested payment and provided payment instructions to me (XXXX). However, when Ms. XXXX contacted Bank of America, which was evidently just after the aforementioned email, she was told to close escrow. In her words, quoted from an email to me sent one hour after the previous one (XXXX XXXX, XXXX XXXX) : " I will contact Bank of America, to verify some info and they informed me that they will be escrowing the insurance, so they asked me to forward the last email with the payment instructions. " Apparently, Ms. XXXX failed to properly inform her insurance department and they went ahead with the escrow payment, contrary to my request and in violation of my agreement with Ms. XXXX. \n\nAfter payment was cancelled, but Ms. XXXX obstructed my attempts to settle the matter, and rather than risk making a payment to my insurance agent, I sent an email (dated Fri. XXXX XXXX) to my insurance agent Ms. XXXX : [beginning of quoted email] XXXX recommendation, I am putting the three of us in conversation. \n\nAt this point, here is the resolution I would like to propose: BofA to XXXX/XXXX of XXXX should proceed for this year's insurance premium. (XXXX, please do not proceed with the credit premium should not be made from escrow -- please bill me directly. \n\nMs. XXXX, please cancel the pending property tax payment noted on the escrow account), and instruct the relevant departments not to make any future payments from that account. \n\nPlease note on the escrow account, please close it immediately. \n\nPlease let me know whether this solution is possible. \n\nThank you for receiving a reply to this email. My insurance agent did as requested, but Bank of America not only failed to remove the upcoming tax bill, but has increased my escrow payment, as of the mortgage bill due XXXX XXXX. \n\nRegarding the tax issue, it appears that his matter have been obstructed by Ms. XXXX and Bank of America. Previously, during the conversation between myself and Ms. XXXX, she cited the Bank of America as the reason why the tax payment could not be removed from the escrow account. At that time, my request was refused -- Ms. XXXX cited the Bank of America as the reason why the tax payment could not be removed from the escrow account. However, as I explained on XXXX XXXX, Bank of America is responsible for my city tax as XXXX XXXX, XXXX. In fact, property taxes are and have always been due in the City of XXXX XXXX on April 1st and sent out, not the due date. The website of the XXXX XXXX City Collector of Revenue leaves no doubt: " Each XXXX, the City sends a notice to all landowners. Taxes are due by XX/XX/XXXX of each year " (XXXX XXXX XXXX). \n\nIndeed, Bank of America is a fact that in past years tax payments made by Bank of America from escrow on my behalf were dated XXXX XXXX (XXXX) and XXXX. I informed this to Ms. XXXX on XXXX XXXX and she agreed that she would see to it that the tax payment was removed, but she would not do so unless I paid the amount due in full. \n\nI am currently in the process of removing the tax payment from the escrow account, but I am still awaiting confirmation from Ms. XXXX that the payment has been removed. \n\nI hope this information is helpful. \n\nThank you for your attention to this matter.

> docs # Check the corpus

<<VCorpus>>

Metadata: corpus specific: 0, document level (indexed): 1

Content: documents: 27536

Corpus

Text Pre-processing

DTM

```
> lapply(docs, content)[c(10, 100, 1000)] # Check the content
$ 10
[1] "I have been requesting Bank of America remove PMI from my account for over 5 years. \n\nI w
t the letter stated Bank of America needs a property valuation with a fee of {$100.00}. \n\nI pa
unt of {$100.00}. On XX/XX/2023 I called Bank of America about PMI removal and spoke to someone
To my surprise He advised that regardless of what the original letter stated it was no longer val
0 days. \n\nWhere in that original letter does it state This letter is valid for 30 days. Or
no later than 30 days. \n\nThis is unfair to consumers and should have been disclosed. It appear
once again. \n\nHow much money is Bank of America making by holding these PMI payments for years
nd delays to avoid deleting PMI."
$ 100
[1] "We are in a Hurricane Irma FEMA designated area ( XXXX Florida- XXXX , FL vicinity ). \n\nW
eferred minimum payment grace period on our Bank of America credit card ending in -XXXX. \n\nWe
ent as promised in XX/XX/XXXX until we filed a CFPB complaint in XX/XX/XXXX. Bank of America I
elp us resolve the issue for deferred minimum payment. \n\nWhen the deferred minimum payment iss
a, it was done very poorly. \n\nBank of America has been the LEAST helpful to Hurricane Irma vic
e can name a dozen other banks who have been so much more understanding and helpful. Bank of Ame
account hurricane fees of {$27.00} on XX/XX/XXXX ; {$38.00} on XX/XX/XXXX ; {$38.00} on XX/XX/XXXX
as caused us great emotional distress by reporting to the credit bureaus a late payment notificati
ment until Hurricane Irma. Our local TV station is requesting to do an in-depth expose on the tr
ed Hurricane Irma victims in our area, which has been horrible. \n\nWe should not be charged rid
on XX/XX/XXXX ; {$38.00} on XX/XX/XXXX ; {$38.00} on XX/XX/XXXX nor be subjected to a credit bur
to Hurricane Irma. We feel like Bank of America has no compassion for Hurricane Irma victims, es
category. \n\nXXXX XXXX"
$ 1000
[1] "XXXX % XXXX check could not be processed by a XXXX insurance carrier. The Check has the c
ansit & Account number along with Item number encoded. \nOn the first failed attempt BofA validat
ut could not locate an attempt for the item to clear the bank. \n\nAfter XXXX attempts to process
of America, I authorized the transaction using the card as the medium. \n\nIn a followup call, B
by transferring the card based transaction into the promotion. \n\nIn short, this scheme of prov
itutes a bait & switch in an attempt to defraud the customer. \n\nI requested he log a customer
ich he committed to do, but could not ( or would not ) tell me if I would be contracted in resp
ed him that I would make this regulator complaint to be sure the complaint was heard an acted on
```



```
> lapply(docs, content)[c(10, 100, 1000)] # Check the content
$ 10
[1] "request bank america remove pmi account account eligible letter bank america property va
america pmi removal speak pmi department surprise advise original letter statedit valid day o
mer send payment late day unfair consumer disclose process delay money bank america hold pmi p
lete pmi"
$ 100
[1] "hurricane irma designate florida fl vicinity apply grant defer minimum payment grace
ive payment deferment promise file cfpb complaint bank america irma relief resolve issue defe
nt issue correct bank america bank america helpful hurricane irma victim bank institution do
ica charge credit card account hurricane fee additionally bank america emotional distress repe
cation late payment hurricane irma local station request indepth expose treatment bank americ
ble charge ridiculous hurricane fee subject credit bureau late payment notification hurricane
rricane irma victim minority category"
$ 1000
[1] "check process insurance carrier check customary route transit account numb numb encode fo
e attempt bank attempt process check supply bank america authorize transaction card medium fo
nsfer card base transaction promotion short scheme provide nonnegotiable check constitute be
quest log customer complaint commit contract response complaint inform regulator complaint co
```

Corpus



Text Pre-processing



DTM

```
> lapply(docs, content)[c(10, 100, 1000)] # Check the content
$`10`
[1] "request bank america remove pmi account account eligible letter bank america property va
america pmi removal speak pmi department surprise advise original letter statedit valid day of
mer send payment late day unfair consumer disclose process delay money bank america hold pmi p
lete pmi"

$`100`
[1] "hurricane irma fema designate florida fl vicinity apply grant defer minimum payment grace
eive payment deferment promise file cfpb complaint bank america irma relief resolve issue def
ent issue correct bank america bank america helpful hurricane irma victim bank institution doz
ica charge credit card account hurricane fee additionally bank america emotional distress rep
cation late payment hurricane irma local station request indepth expose treatment bank americ
ble charge ridiculous hurricane fee subject credit bureau late payment notification hurricane
ricane irma victim minority category"

$`1000`
[1] "check process insurance carrier check customary route transit account numb numb encode fo
e attempt bank attempt process check supply bank america authorize transaction card medium fo
nsfer card base transaction promotion short scheme provide nonnegotiable check constitute be
quest log customer commit contract response complaint inform regulator complaint co
```

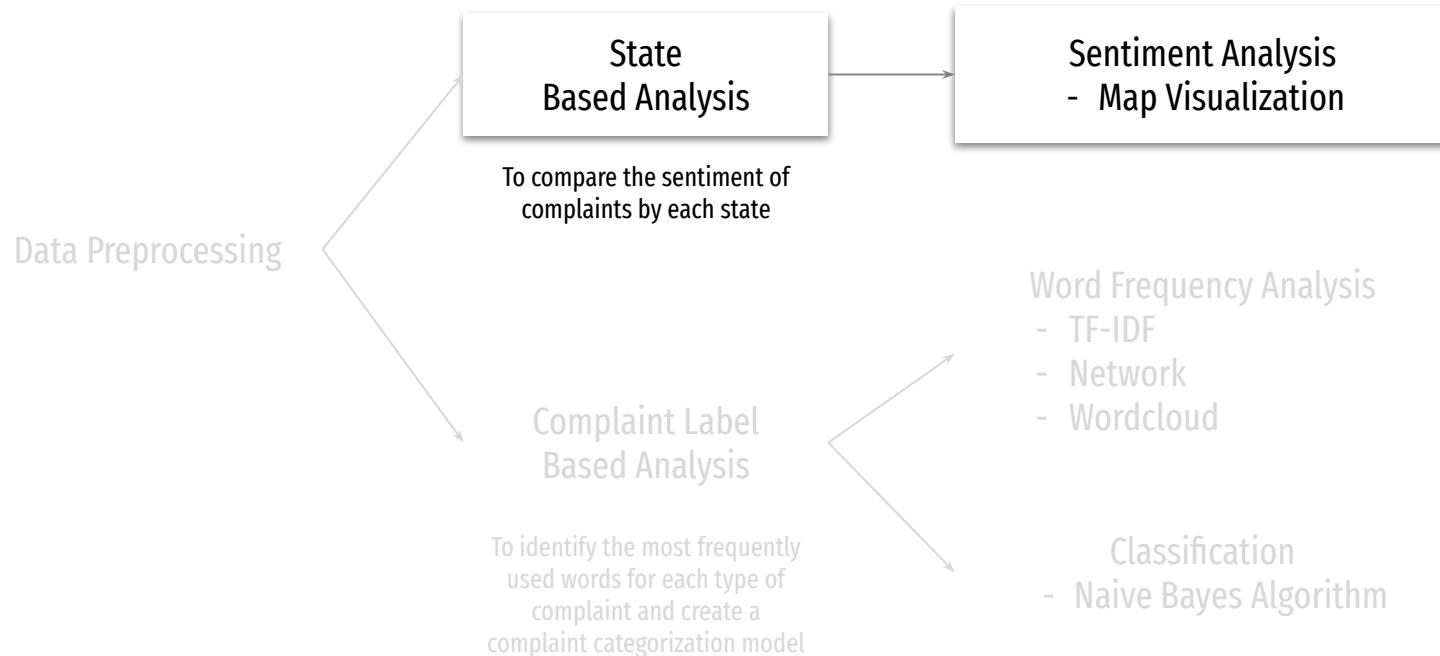


```
> inspect(dtm) # Check the DTM
<<DocumentTermMatrix (documents: 27536, terms: 703)>>
Non-/sparse entries: 912738/18445070
Sparsity : 95%
Maximal term length: 14
Weighting : term frequency (tf)
Sample : 

Terms
Docs account america bank card check credit money payment receive report
10068 35 1 1 0 1 98 0 28 0 116
10807 8 5 10 0 1 2 4 1 1 1
13893 30 15 38 0 57 0 1 4 18 5
15010 35 1 1 0 1 98 0 28 0 116
20395 19 0 7 0 0 19 1 7 6 75
21676 11 1 1 1 1 68 0 16 0 102
24375 14 20 36 1 1 2 4 13 1 4
26884 23 0 1 0 3 29 1 3 3 73
5283 6 1 1 0 1 53 0 11 0 82
5513 15 1 1 0 1 74 0 18 0 109
```

- Limit the word length (3, Inf)
- Lower (0.05%) and upper limits (95%) for the number of documents that appear
- Remove rows with sparsity over 0.99

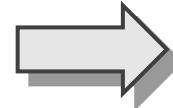
Analysis Overview



Preprocessing Steps

Use sentiment method 'afinn'

```
> State.affin %>% head()  
# A tibble: 6 × 3  
  State word      value  
  <chr> <chr>    <dbl>  
1 AE    fraudulent -4  
2 AE    charges    -2  
3 AE    no         -1  
4 AE    no         -1  
5 AE    restricted -2  
6 AE    dire       -3
```



Sum of each group value

```
> total_State_sum %>% head()  
# A tibble: 6 × 2  
  State value  
  <chr> <dbl>  
1 AE    -43  
2 AK    -157  
3 AL    -190  
4 AP    -2  
5 AR    -56  
6 AZ    -954
```

Standardization

```
> total_State_sum %>% head()
```

A tibble: 6 × 4

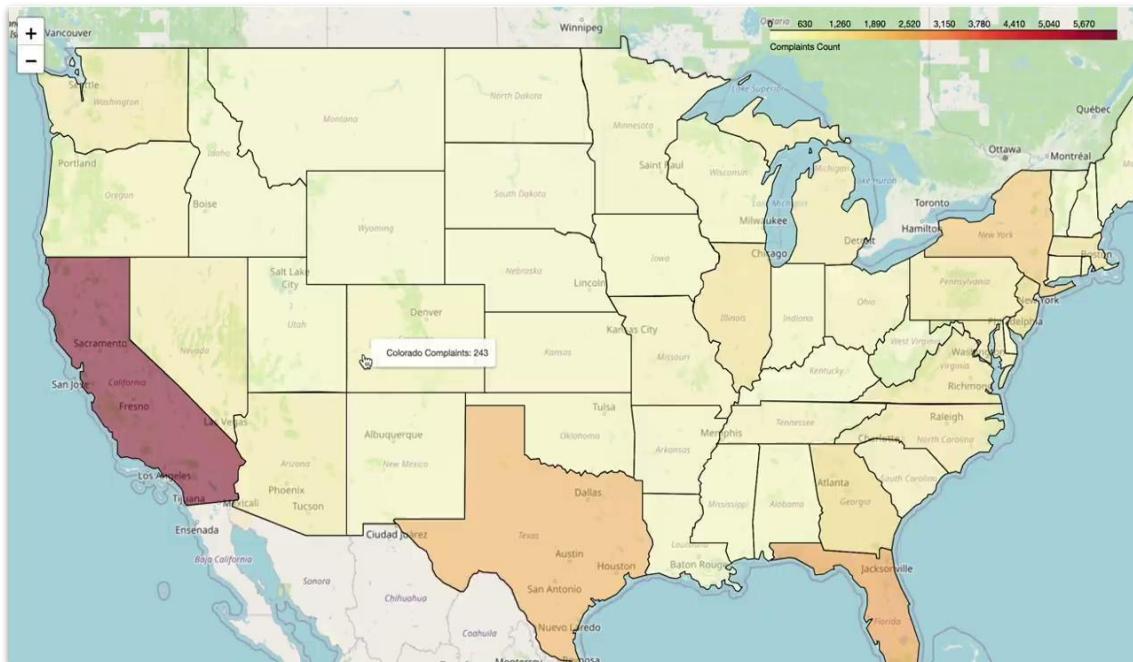
	State	value	count	pro
1	AE	-43	6	-7.17
2	AK	-157	14	-11.2
3	AL	-190	34	-5.59
4	AP	-2	2	-1
5	AR	-56	31	-1.81
6	AZ	-954	174	-5.48

Sentiment Score of a State

Complaint Count of a State

To standardize the varying counts across states,
we create a column named 'pro' by dividing the value by the count.

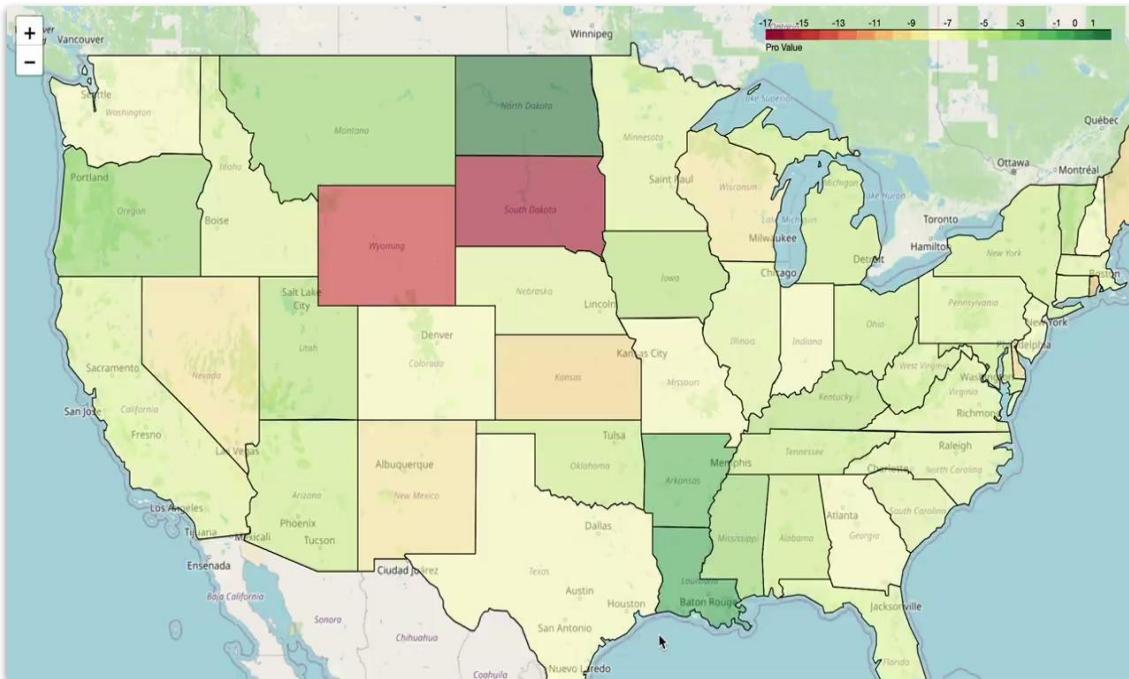
Complaint Visualization



Top 5 States by
Number of Complaints

1. California
2. Florida
3. Texas
4. New York
5. Georgia

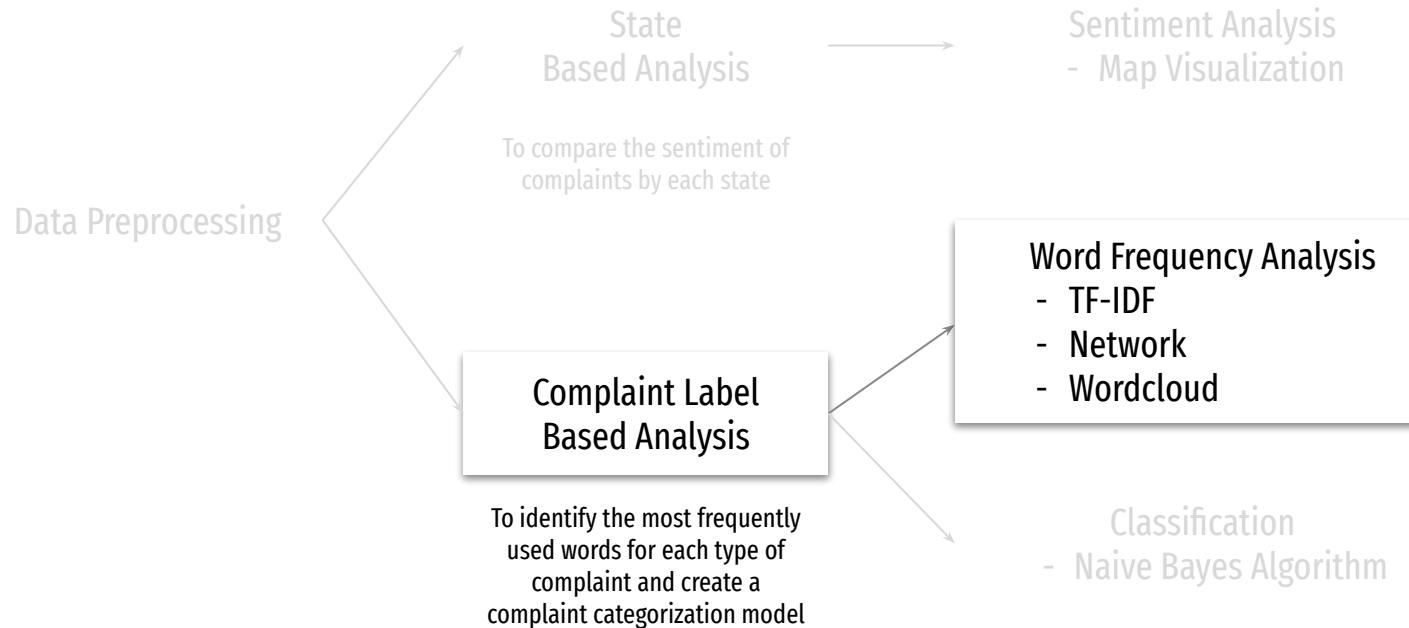
Sentiment Visualization



Top 5 States by
Negative Sentiments

1. South Dakota
2. Hawaii
3. Wyoming
4. Alaska
5. Rhode Island

Analysis Overview



Top 5 TF-IDF words

```
> comp.table %>% arrange(desc(TF)) %>% head(10)
```

	doc	term	TF	TF_IDF	W_TF_IDF
1	1	bank	72151	0	0
2	1	account	59281	0	0
3	1	america	44265	0	0
4	1	card	28405	0	0
5	1	check	23326	0	0
6	1	credit	23046	0	0
7	1	money	18978	0	0
8	1	claim	17623	0	0
9	1	receive	17414	0	0
10	1	charge	17022	0	0

```
> comp.table %>% arrange(desc(TF_IDF)) %>% head(10)
```

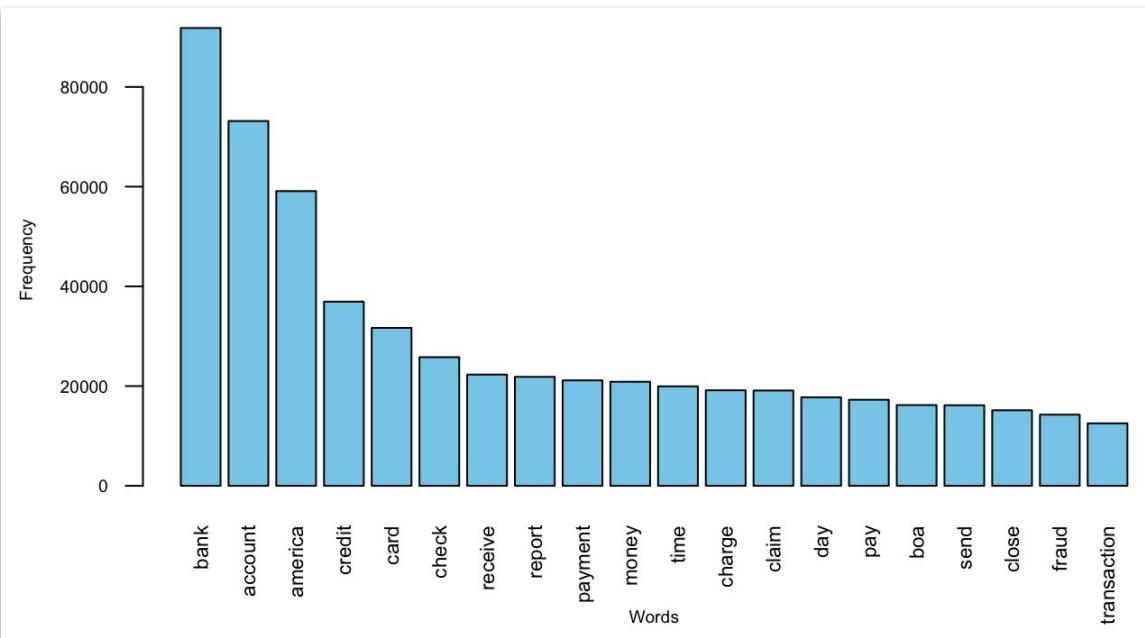
	doc	term	TF	TF_IDF	W_TF_IDF
1	3	pmi	218	345.5218	0.001
2	1	merchandise	371	217.0211	0.000
3	1	flight	349	204.1519	0.000
4	1	fraudster	306	178.9985	0.000
5	1	chip	276	161.4497	0.000
6	1	puppy	94	148.9865	0.000
7	1	cellphone	81	128.3820	0.000
8	1	harmless	81	128.3820	0.000
9	1	mastercard	198	115.8226	0.000
10	1	redeposit	64	101.4376	0.000

↑ High TF does not guarantee high TF-IDF ↑

```
> top_terms_df # Print the result (Top 5 TF-IDF words in each document)
```

	Doc_id	Term1	Term2	Term3	Term4	Term5
Card&Account	1	merchandise	flight	fraudster	chip	puppy
Credit_management	2	subscriber	antedate	creditreporting	permissible	furnishers
Loan	3	pmi	ltv	haf	recontrust	recast

Top 20 Words



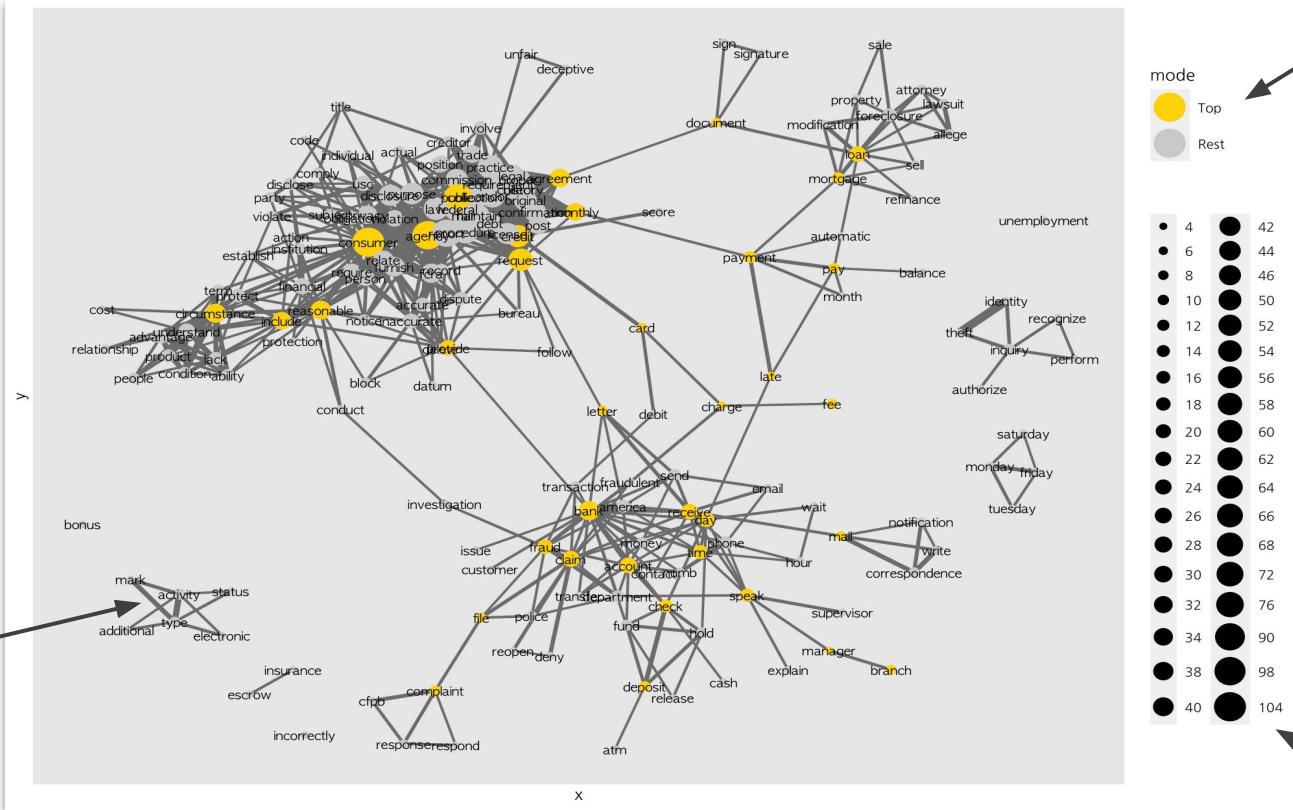
Top 5 Words by Word Frequency

1. bank
2. account
3. america
4. credit
5. card

Network Analysis

Vertices = 703

Edges = 544

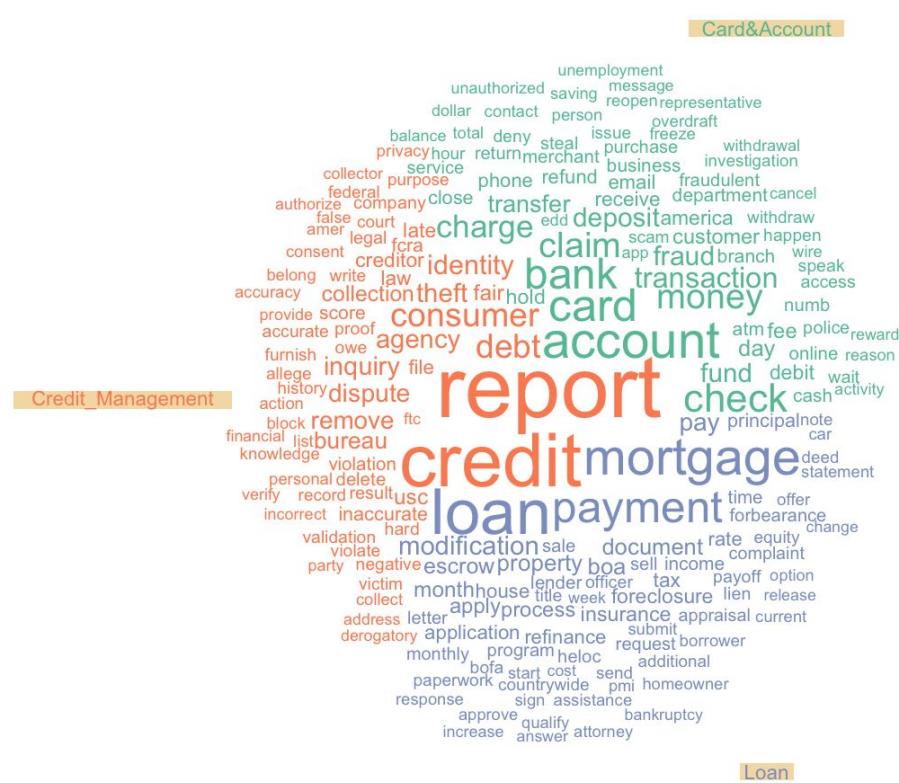


Betweenness
Top 5%

Degree

Closeness

Comparison Wordcloud

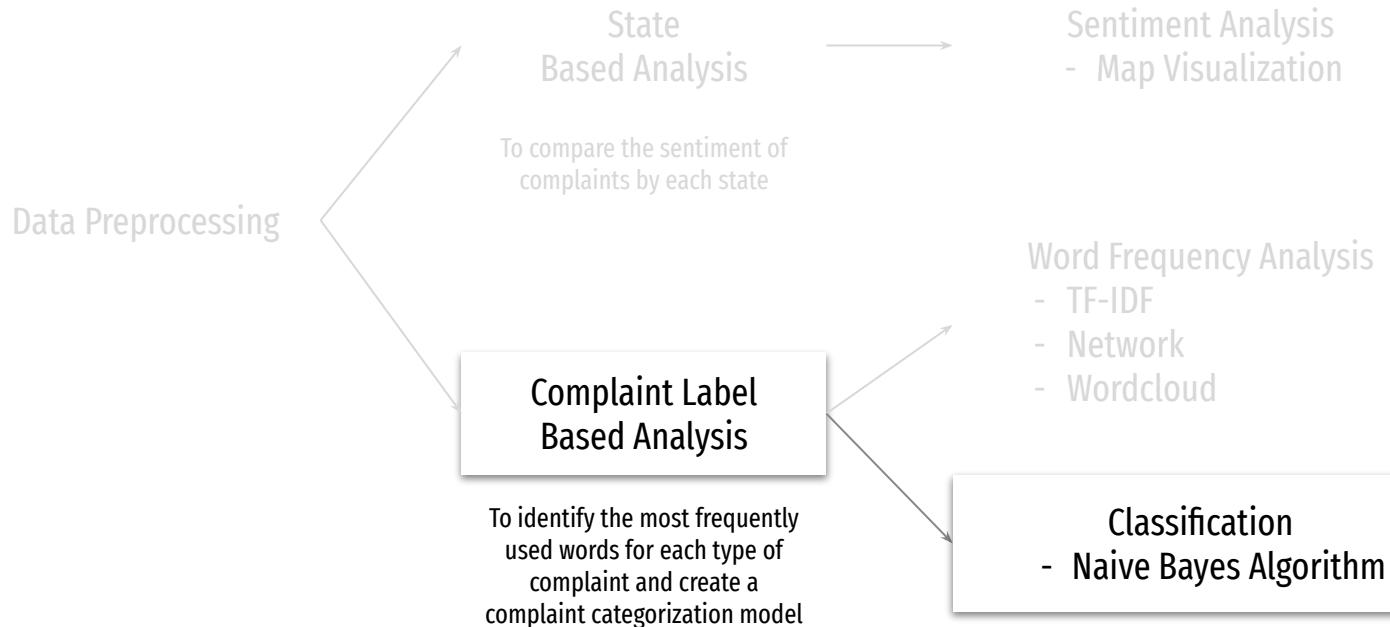


	aaa	abandon	abide	ability	abnormal	abovementioned
Card&Account	56	37	60	387	20	12
Credit_Management	23	8	31	139	1	5
Loan	0	15	12	70	0	0

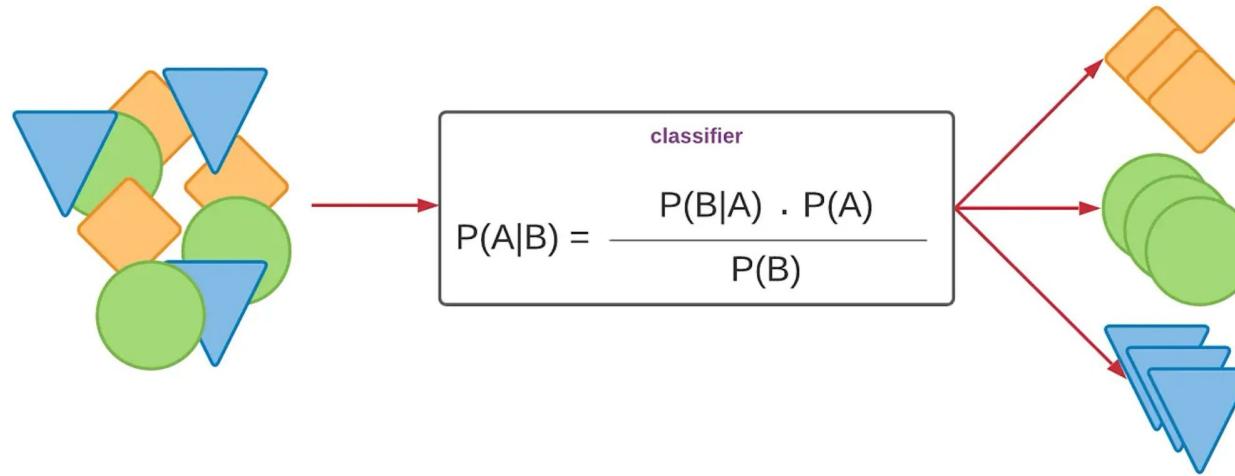
```
set.seed(123)
```

```
# Create a comparison cloud by transposing the matrix
# Rows become words, columns become categories
comparison.cloud(t(comparison.label),
  # Title size
  title.size = 0.7,
  # Color palette from Set2 with 3 colors
  colors = brewer.pal(3, "Set2"),
  # Title color palette
  title.colors = brewer.pal(3, "Set2"),
  # Random order False
  random.order=FALSE,
  # Title background color
  title.bg.colors = "wheat",
  # Rotation percentage (0 for horizontal)
  rot.per = 0,
  # Scale for word sizes
  scale = c(3, 0.5),
  # Maximum number of words
  max.words = 200)
```

Analysis Overview

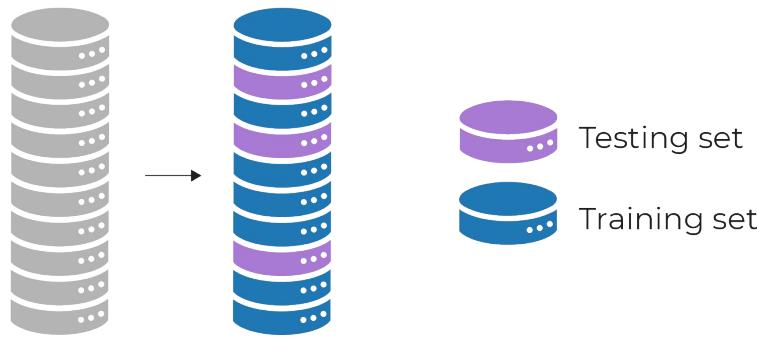


Naive Bayes Classification



Naive Bayes classification algorithm is a probabilistic machine learning method based on Bayes' theorem, assuming independence between features, and is particularly effective for text classification tasks.

Train & Test Set Split

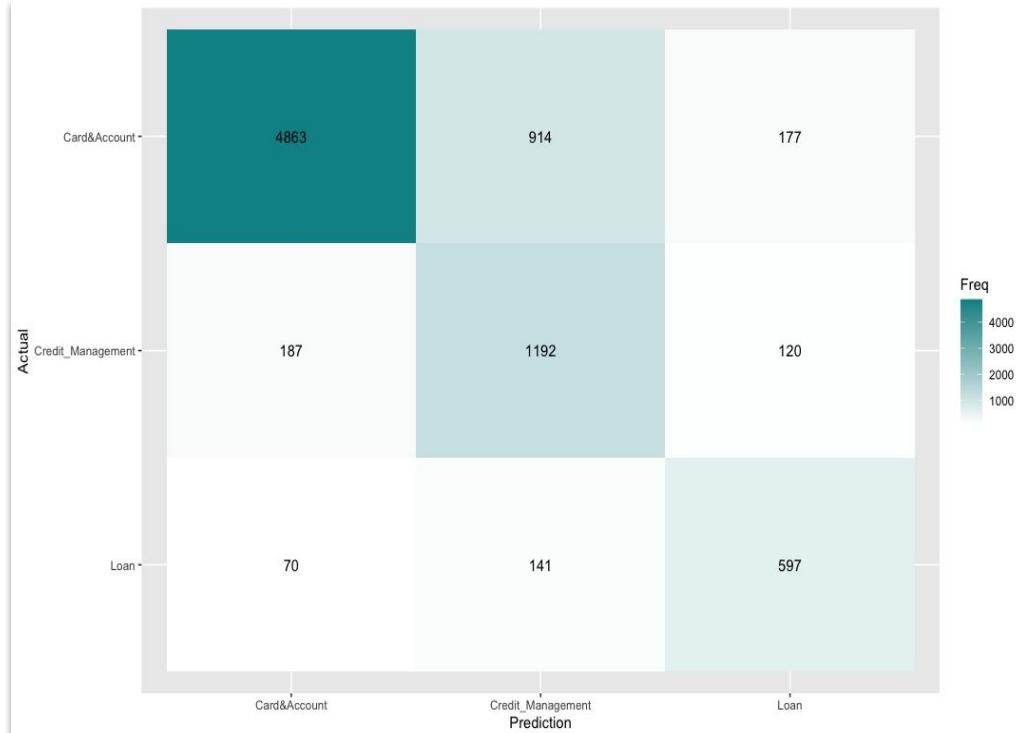


```
> table(y.train) # Number of labels in train data  
y.train  
Card&Account Credit_Management      Loan  
    14002            3455        1818  
> table(y.test) # Number of labels in test data  
y.test  
Card&Account Credit_Management      Loan  
    5995            1451        815
```

Train and test set well reflect the proportions of the entire dataset.

```
> prop.table(table(complaint$type)) # Entire dataset  
  
Card&Account Credit_Management      Loan  
    0.72621296      0.17816676      0.09562028  
> prop.table(table(y.train)) # Train dataset  
y.train  
Card&Account Credit_Management      Loan  
    0.72643320      0.17924773      0.09431907  
> prop.table(table(y.test)) # Test dataset  
y.test  
Card&Account Credit_Management      Loan  
    0.72569907      0.17564460      0.09865634
```

Accuracy of the Model



$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

```
> evaluate_model(y.test, complaint.nb.pred)
Accuracy: 0.81
Precision: 0.53
Recall: 0.8
F1 Score: 0.64
```

CONCLUSION



States Analysis

The investigation of state-wise complaints revealed that California had the highest number of grievances filed, but sentiment analysis showed that the severity of dissatisfaction was greatest in South Dakota.

This suggests the importance of managing states with a high volume of complaints, and also highlights that the depth of dissatisfaction is a crucial factor in consumer grievance management. It confirms the need to consider both these aspects thoroughly.

Label Analysis & Classification

We reclassified the existing categories by grouping similar types together. As a result, when a complaint is filed, it can be quickly categorized into one of the three broad categories we established, allowing for proactive issue resolution and classification.

In practice, this method showed that similar words appeared within the categories we grouped together, and the model demonstrated a accuracy of 81%.

REFERENCES

- Text Mining Lecture Slides
- [Problem Definition]
<https://punchng.com/banks-record-117-increase-in-customers-complaints/>
- [Network]
https://bookdown.org/yuaye_kt/RTIPS/Texnetwork-2.html#%EB%84%A4%ED%8A%B8%EC%9B%8C%ED%81%AC-%EA%B0%9D%EC%B2%B4-%EC%B6%94%EC%B6%9C



THANK YOU