

```

import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
import re
from bs4 import BeautifulSoup
from tqdm import tqdm
from nltk.stem import WordNetLemmatizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, f1_score, confusion_matrix
pd.set_option("display.max_colwidth", None)

```

```
complaints = pd.read_csv(r"C:\Users\sinha\Bank complaint.csv")
```

```
complaints.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 18 columns):

```

#	Column	Non-Null Count	Dtype
0	Date received	1048575 non-null	object
1	Product	1048575 non-null	object
2	Sub-product	874198 non-null	object
3	Issue	1048575 non-null	object
4	Sub-issue	699655 non-null	object
5	Consumer complaint narrative	383564 non-null	object
6	Company public response	449051 non-null	object
7	Company	1048575 non-null	object
8	State	1031078 non-null	object
9	ZIP code	935181 non-null	object
10	Tags	146542 non-null	object
11	Consumer consent provided?	690654 non-null	object
12	Submitted via	1048575 non-null	object
13	Date sent to company	1048575 non-null	object
14	Company response to consumer	1048568 non-null	object
15	Timely response?	1048575 non-null	object
16	Consumer disputed?	534721 non-null	object
17	Complaint ID	1048575 non-null	int64

```
dtypes: int64(1), object(17)
```

```
memory usage: 144.0+ MB
```

```
complaints.columns
```

```

Index(['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'],
      dtype='object')
```

```
complaints["Consumer complaint narrative"]
```

```
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
```

```
...
1048570 NaN
1048571 NaN
1048572 NaN
1048573 NaN
1048574 NaN
```

```
Name: Consumer complaint narrative, Length: 1048575, dtype: object
```

```
complaints=complaints[~pd.isnull(complaints['Consumer complaint
narrative'])]
```

```
complaints.head()
```

```
      Date received \
29904    03/23/2019
30629    03/22/2019
30735    03/22/2019
30795    03/22/2019
30807    03/22/2019
```

```
Product \
29904  Credit reporting, credit repair services, or other personal
consumer reports
30629  Credit reporting, credit repair services, or other personal
consumer reports
30735  Credit reporting, credit repair services, or other personal
consumer reports
30795  Credit reporting, credit repair services, or other personal
consumer reports
30807  Credit reporting, credit repair services, or other personal
consumer reports
```

```
      Sub-product      Issue \
29904  Credit reporting  Incorrect information on your report
30629  Credit reporting  Incorrect information on your report
30735  Credit reporting  Incorrect information on your report
```

30795 Credit reporting Incorrect information on your report
30807 Credit reporting Incorrect information on your report

Sub-issue \
29904 Account information incorrect
30629 Account information incorrect
30735 Account information incorrect
30795 Account information incorrect
30807 Account information incorrect

Consumer complaint narrative \

29904 The Summer of XX/XX/2018 I was denied a mortgage loan due to a charge off from XXXX XXXX credit card. I both mailed an account validation letter and disputed this debt with the credit bureaus and the credit card company several times. Only the credit bureaus responded that this debt was verified but showed no proof as I requested. XXXX XXXX still has not responded to any of my letters since XX/XX/2018. All I asked was that the credit bureaus show me how they verified this debt and what XXXX XXXX sent them as proof of my debt other than a bill that could be fabricated. I really need to know why I haven't been shown proof of this after multiple disputes and certified letters letters sent.

30629

There are many mistakes appear in my report without my understanding.

30735

There are many mistakes appear in my report without my understanding.

30795

There are many mistakes appear in my report without my understanding.

30807

There are many mistakes appear in my report without my understanding.

Company public response \

29904 Company has responded to the consumer and the CFPB and chooses not to provide a public response
30629 Company has responded to the consumer and the CFPB and chooses not to provide a public response
30735 Company has responded to the consumer and the CFPB and chooses not to provide a public response
30795 Company has responded to the consumer and the CFPB and chooses not to provide a public response
30807 Company has responded to the consumer and the CFPB and chooses not to provide a public response

Company State ZIP code Tags \

29904	TRANSUNION INTERMEDIATE HOLDINGS, INC.	IL	NaN	NaN
30629	TRANSUNION INTERMEDIATE HOLDINGS, INC.	VA	220XX	NaN
30735	TRANSUNION INTERMEDIATE HOLDINGS, INC.	TX	770XX	NaN
30795	TRANSUNION INTERMEDIATE HOLDINGS, INC.	TX	787XX	NaN
30807	TRANSUNION INTERMEDIATE HOLDINGS, INC.	CA	951XX	NaN

	Consumer consent provided?	Submitted via	Date sent to company	\
29904	Consent provided	Web	03/23/2019	
30629	Consent provided	Web	03/22/2019	
30735	Consent provided	Web	03/22/2019	
30795	Consent provided	Web	03/22/2019	
30807	Consent provided	Web	03/22/2019	

	Company response to consumer	Timely response?	Consumer disputed?
29904	Closed with explanation	Yes	NaN
30629	Closed with explanation	Yes	NaN
30735	Closed with explanation	Yes	NaN
30795	Closed with explanation	Yes	NaN
30807	Closed with explanation	Yes	NaN

	Complaint ID
29904	3189109
30629	3187982
30735	3187954
30795	3188091
30807	3188119

complaints.shape

(383564, 18)

complaints["Product"].value_counts()

Credit reporting, credit repair services, or other personal consumer reports 92378

Debt collection

86710

Mortgage

52987

Credit reporting

31588

Student loan

21810

Credit card or prepaid card

21379

```

Credit card
18838
Bank account or service
14885
Checking or savings account
12881
Consumer Loan
9474
Vehicle loan or lease
5745
Money transfer, virtual currency, or money service
5466
Payday loan, title loan, or personal loan
4421
Payday loan
1747
Money transfers
1497
Prepaid card
1450
Other financial service
292
Virtual currency
16
Name: Product, dtype: int64

complaints["Product"].nunique()

18

complaints['Product'] = complaints['Product'].replace(['Student
loan','Vehicle loan or lease',
                                                         'Consumer Loan','Payday loan,
title loan, or personal loan', 'Payday loan'], 'Loan')
complaints['Product'] = complaints['Product'].replace(["Bank account
or service","Checking or savings account"], 'Bank_Accounts')

complaints['Product'] = complaints['Product'].replace(["Money
transfers","Money transfer, virtual currency, or money service",
                                                         "Virtual currency","Other
financial service"], 'Money_Service_or_Currency_Service')
complaints['Product'] = complaints['Product'].replace(["Prepaid
card","Credit card or prepaid card","Credit card"], 'cards')
complaints['Product'] = complaints['Product'].replace(["Credit
reporting, credit repair services, or other personal consumer
reports",
                                                         "Credit reporting"],,
'Credit_Reporting')
complaints['Product'] = complaints['Product'].replace(["Debt
collection"], 'Debt_collection')

```

```
complaints['Product'].value_counts()
```

Credit_Reporting	123966
Debt_collection	86710
Mortgage	52987
Loan	43197
cards	41667
Bank_Accounts	27766
Money_Service_or_Currency_Service	7271

Name: Product, dtype: int64

```
# Mortgage-
# Debt collection-
# Loans -
# Student loan,
# Consumer Loan,
# vehicle loan or lease,
# Payday loan, title loan, or personal loan
# Payday loan

# Cards-
# Credit card
# Credit card or prepaid card
# Prepaid card

# Credit reporting-
# Credit reporting, credit repair services, or other personal consumer
reports
# Credit reporting

# vehicle loan or lease,
# Payday loan, title loan, or personal loan
# Payday loan

# Cards-
# Credit card
# Credit card or prepaid card
# Prepaid card

# Credit reporting-
# Credit reporting, credit repair services, or other personal consumer
reports
# Credit reporting

# Bank account services-
# Bank account or service
# Checking or savings account
# Other financial service

# Currency Services-
```

```
# Money transfer, virtual currency, or money services  
# Money transfers  
# Virtual currency
```

```
complaints.shape
```

```
(383564, 18)
```

```
complaints_train,complaints_test=train_test_split(complaints,test_size  
= 0.5, random_state = 42)
```

```
complaints_train.shape
```

```
(191782, 18)
```

```
complaints_test.shape
```

```
(191782, 18)
```

```
def decontract(text):
```

```
    text = re.sub(r"won't", "will not", text)
```

```
    text = re.sub(r"can't", "can not", text)
```

```
    text = re.sub(r"n't", " not", text)
```

```
    text = re.sub(r"\ 're", " are", text)
```

```
    text = re.sub(r"\ 's", " is", text)
```

```
    text = re.sub(r"\ 'd", " would", text)
```

```
    text = re.sub(r"\ 'll", " will", text)
```

```
    text = re.sub(r"\ 't", " not", text)
```

```
    text = re.sub(r"\ 've", " have", text)
```

```
    text = re.sub(r"\ 'm", " am", text)
```

```
    return text
```

```
stop_words= set(['br', 'the', 'i', 'me', 'my', 'myself', 'we', 'our',  
'ours', 'ourselves', 'you', "you're", "you've",\  
                "you'll", "you'd", 'your', 'yours', 'yourself',\  
'yourselves', 'he', 'him', 'his', 'himself', \  
                'she', "she's", 'her', 'hers', 'herself', 'it', "it's",  
'its', 'itself', 'they', 'them', 'their',\  
                'theirs', 'themselves', 'what', 'which', 'who', 'whom',  
'this', 'that', "that'll", 'these', 'those', \  
                'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being',  
'have', 'has', 'had', 'having', 'do', 'does', \  
                'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if',  
'or', 'because', 'as', 'until', 'while', 'of', \  
                'at', 'by', 'for', 'with', 'about', 'against', 'between',  
'into', 'through', 'during', 'before', 'after',\  
                'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out',  
'on', 'off', 'over', 'under', 'again', 'further',\  
                'then', 'once', 'here', 'there', 'when', 'where', 'why',
```

```
'how', 'all', 'any', 'both', 'each', 'few', 'more', \
    'most', 'other', 'some', 'such', 'only', 'own', 'same',
'so', 'than', 'too', 'very', \
    's', 't', 'can', 'will', 'just', 'don', "don't", 'should',
"should've", 'now', 'd', 'll', 'm', 'o', 're', \
    've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't",
'didn', "didn't", 'doesn', "doesn't", 'hadn', \
    'hadn't', 'hasn', "hasn't", 'haven', "haven't", 'isn',
"isn't", 'ma', 'mightn', "mightn't", 'mustn', \
    'mustn't', 'needn', "needn't", 'shan', "shan't",
'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", \
    'won', "won't", 'wouldn', "wouldn't"])
```

```
lemmatizer = WordNetLemmatizer()
```

```
def preprocess_text(review):
    review = re.sub(r"http\S+", "URLs", review) # removing
website links
    review = BeautifulSoup(review, 'lxml').get_text() # removing
html tags
    review = decontract(review) #
decontracting
    review = re.sub("\S*\d\S*", "", review).strip() # removing the
words with numeric digits
    review = re.sub('[^A-Za-z]+', ' ', review) # removing
non-word characters
    review = review.lower() # converting
to lower case
    review = [word for word in review.split(" ") if not word in
stop_words] # removing stop words
    review = [lemmatizer.lemmatize(token, "v") for token in review]
#Lemmatization
    review = [word for word in review if len(word)>=3] # alok's :
Alok s
    review = " ".join(review)
    review.strip()
    return review
```

```
final_complaints=complaints_train[["Consumer complaint
narrative", "Product"]]
```

```
import time
st=time.time()
final_complaints['Consumer_complaint'] =final_complaints['Consumer
complaint narrative'].apply(lambda x: preprocess_text(x))
et=time.time()
print("Total Time taken", et-st)
```

```
Total Time taken 149.87934350967407
```


150/60

2.5

```
final_complaints.to_csv(r"C:\Users\sinha\Downloads\final_complaints")
```

```
bank_complaints = pd.read_csv(r"C:\Users\sinha\Downloads\  
final_complaints")
```

```
bank_complaints.head()
```

```
   Unnamed: 0  \  
0      789746  
1      763555  
2      586986  
3      526971  
4      536811
```

Consumer complaint narrative \

0 We are filing this Federal Consumer complaint against our mortgage servicer Nationstar Mortgage (NSM) and the " notorious foreclosure mill " known as Law Offices XXXX Missouri XXXX ; with relation to violations of the Missouri Non-judicial foreclosure laws and procedures act. *Chapter XXXX Mo. Rev. Stat. ; XXXX v. XXXX XXXX XXXX XXXX, XXXX XXXX XXXX XXXX (Mo. XX/XX/XXXX XXXX Our complaint in particular surrounds violations to procedures to Notice of sale. It is the responsibility of NSM as well as XXXX to contain the date, book and page of the Deed of Trust, name of each grantor on the Deed of Trust, the time, terms and place of sale, and a description of the real estate to be sold. The notice must be published and mailed to required parties.XXXX registered agents of NSM have further failed to comply with proper notice with relation to publications as follows:1. PublicationPursuant to Section XXXX Mo XXXX., a Notice of Foreclosure Sale must be printed in a newspaper published in the county in which the real estate is situated at least once a week for a total of four weeks, with the last publication being not more than one week before the sale date. If the land is located in more than XXXX XXXX, the notice must be published in each XXXX. If the land is located in a Missouri XXXX XXXX, of which there are XXXX, publication is for XXXX consecutive days in a daily newspaper with the last day being the date of sale.Furthermore, XXXX, registered agents of NSM have further failed to comply with proper notice with relation to mailing as follows with relation to notice to XXXX XXXX recorder 's office via mail ; of the intent to execute foreclosure action on our real property XX/XX/XXXX, at XX/XX/XXXX:2. MailPursuant to section XXXX Mo. Rev. XXXX, the notice of sale must be mailed via certified or registered mail, postage prepaid and deposited in the U.S. mail at least 20 days before the date of the foreclosure sale. Such notice must be sent to the last known address of the original grantors in the

Deed of Trust, the last known address of person 's shown by the recorder 's office to be the owners of the property as of 40 days before the date of the foreclosure sale, and any person who has recorded a request for notice of sale under the Deed of Trust either at least 40 days before the date of the sale or recorded under another Deed of Trust that was recorded before the Deed of Trust now being foreclosed, if that prior Deed of Trust is not released of record. Lastly, there are XXXX liens totaling in the amount of approximately {\$11000.00}, which are outstanding liens on title for our home. Our records and attempts have indicated that the XXXX, has not been notified According to XXXX XXXX XXXX Section XXXX, of the intent to execute foreclosure action on our real property XX/XX/XXXX, at XXXX : 3. Notice to XXXX According to XXXX XXXX XXXX Section XXXX, notice of the foreclosure sale must be sent to the Internal Revenue Service if an XXXX Notice of Lien on the real estate is filed more than 30 days before the date of the sale. This notice must be given by personal service or by registered or certified mail to the XXXX District Director at least 25 days before the date of sale. We implore your office to address this matter at once due to our pending foreclosure sale date of XX/XX/XXXX, at XXXX. This illegal process has been conducted very expeditiously and it is an injustice to allow NSM and XXXX to abuse homeowners in this manner.

1

I have an Open account with XXXX Collection they account was paid in FULL before I purchased my house in XXXX, I then File Bankruptcy and was discharged of my debt in XXXX XXXX. Equifax continues to report it has an Open Collection which it is not I have contacted them XXXX time to get this close with proof of payment bankrupt discharge etc. They still refuse to close the account.

2

i sent a letter cerfied mail to equifax about a collection account on my credit report the apartment complex has me down as owing over XXXX i never lived there i apply for the complex but turned it down becauae i could no longer move to a different state at the time i asked equifax in the letter i sent to send proof of me owing this debt and they have not they keep telling me its verifed and i ask them who verifed it and how they will not send me anyproof i know i did not live there so there is no lease with my singarture on it i tried to reach the company and they went into bankrupcy so i can not reach anyone there and the collection agency is giving me a hard time bout getting it removed it is messing with my future of my housing and need some guidance to help me fight this equifax removed it out of public recored but wont take it out of collection even after the courts verified that i never was ask to goto court on this matter

3

I am filing a complaint against Lanier Collection agency wrote them letter requesting information. After viewing a copy of my credit report, I noticed a collection account placed on my credit report from you in from XX/XX/XXXX-XX/XX/XXXX I am requesting that you allow me to

validate the alleged debt. I am unaware of any outstanding medical bills that I possess, and I am seeking the name and hospital/medical provider to which I owe the debt and a detailed breakdown of the fees that I owe. Additionally, I am allowed under the Health Insurance Portability and Accountability Act (HIPAA) to protect my privacy and medical records from third parties. I did not give permission to any of my current or prior medical providers to release any of my medical information to a third party. I am aware that the HIPAA does allow the release of limited information about me but anything more is to only be revealed with the patient 's authorization. Therefore, my request is twofold -- validation of debt and HIPAA authorization. Please provide breakdown of fees including any collection costs and medical charges. Provide a copy of my signature with the provider of service to release my medical information to you. Cease any credit bureau reporting until the debt has been validated by me. They sent me back XXXX pages with total charge and a description the procedure I was supposedly had gotten. They refused to send me anything with my signature provide them service to release any of medical information which is a violation of HIPAA. I never agreed to pay them anything nor do I know who they are. I am unaware of these bills ; these are not my accounts. Then I tried to call to inform them they may have the wrong person. They were extremely rude.

4

Loan Servicing company trying to collect discharged class XXXX bankruptcy debt. see QWR notice of error within.

	Product \
0	Mortgage
1	Credit_Reporting
2	Credit_Reporting
3	Credit_Reporting
4	Mortgage

Consumer_complaint

0 file federal consumer complaint mortgage servicer nationstar mortgage nsm notorious foreclosure mill know law offices xxxx missouri xxxx relation violations missouri non judicial foreclosure laws procedures act chapter xxxx rev stat xxxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx complaint particular surround violations procedures notice sale responsibility nsm well xxxx contain date book page deed trust name grantor deed trust time term place sale description real estate sell notice must publish mail require party xxxx register agents nsm fail comply proper notice relation publications publication pursuant section xxxx xxxx notice foreclosure sale must print newspaper publish county real estate situate least week total four weeks last publication not one week sale date land locate xxxx xxxx notice must publish xxxx land locate missouri xxxx xxxx xxxx publication xxxx consecutive days daily newspaper last day

date sale furthermore xxxx register agents nsm fail comply proper notice relation mail follow relation notice xxxx xxxx recorder office via mail intent execute foreclosure action real property xxxx mailpursuant section xxxx rev xxxx notice sale must mail via certify register mail postage prepay deposit mail least days date foreclosure sale notice must send last know address original grantors deed trust last know address person show recorder office owners property days date foreclosure sale person record request notice sale deed trust either least days date sale record another deed trust record deed trust foreclose prior deed trust not release record lastly xxxx liens total amount approximately outstanding liens title home record attempt indicate xxxx not notify accord xxxx xxxx xxxx section xxxx intent execute foreclosure action real property xxxx xxxx notice xxxxaccording xxxx xxxx xxxx section xxxx notice foreclosure sale must send internal revenue service xxxx notice lien real estate file days date sale notice must give personal service register certify mail xxxx district director least days date sale implore office address matter due pending foreclosure sale date xxxx xxxx illegal process conduct expeditiously injustice allow nsm xxxx abuse homeowners manner

1

open account xxxx collection account pay full purchase house xxxx file bankruptcy discharge debt xxxx xxxx equifax continue report open collection not contact xxxx time get close proof payment bankrupt discharge etc still refuse close account

2

send letter cerfied mail equifax collection account credit report apartment complex owe xxxx never live apply complex turn becauae could longer move different state time ask equifax letter send send proof owe debt not keep tell verifed ask verifed not send anyproof know not live lease singarture try reach company bankrupcy not reach anyone collection agency give hard time bout get remove mess future house need guidance help fight equifax remove public recored wont take collection even court verify never ask goto court matter

3

file complaint lanier collection agency write letter request information view copy credit report notice collection account place credit report xxxx xxxx request allow validate allege debt unaware outstanding medical bill possess seek name hospital medical provider owe debt detail breakdown fee owe additionally allow health insurance portability accountability act hipaa protect privacy medical record third party not give permission current prior medical providers release medical information third party aware hipaa allow release limit information anything reveal patient authorization therefore request twofold validation debt hipaa authorization please provide breakdown fee include collection cost medical charge provide copy signature provider service release medical information cease credit bureau report debt validate send back xxxx page total charge description procedure supposedly get refuse send anything signature provide service release medical information violation hipa never

```

agree pay anything nor know unaware bill not account try call inform
may wrong person extremely rude
4
loan service company try collect discharge class xxxx bankruptcy debt
see qwr notice error within

bank_complaints.shape

(191782, 4)

X=bank_complaints['Consumer_complaint']
y=bank_complaints["Product"]
train_x, val_x, train_y, val_y= train_test_split(X,y ,test_size = 0.5,
random_state = 42)
print("Training data size : ", train_x.shape)

Training data size :  (95891,)

```

Countvectorizer

```

from sklearn.feature_extraction.text import CountVectorizer

vect=CountVectorizer(lowercase=True,
                    stop_words="english",
                    max_df=.4,
                    min_df=1000, ngram_range=(1, 2) )
# min_df= """"Prog will accept only those terms/ words which
# will come into atleast 1000 records""""
# max_df="""" accept only those terms/words which will come into at
# max 20% of documents""""
# ngram_range=(1, 1) How many word we should take together to form a
token

vect.fit(train_x)

CountVectorizer(max_df=0.4, min_df=1000, ngram_range=(1, 2),
               stop_words='english')

vect.get_feature_names()

['ability',
 'able',
 'absolutely',
 'accept',
 'access',
 'accord',
 'account account',
 'account balance',
 'account bank',
 'account charge',
 'account close',

```

'account credit',
'account information',
'account list',
'account number',
'account open',
'account pay',
'account remove',
'account report',
'account tell',
'account xxxx',
'accrue',
'acct',
'accuracy',
'accurate',
'acknowledge',
'act',
'action',
'active',
'activity',
'actual',
'actually',
'add',
'addition',
'additional',
'additionally',
'address',
'address xxxx',
'admit',
'advance',
'advantage',
'advise',
'affect',
'affect credit',
'affidavit',
'afford',
'age',
'agencies',
'agency',
'agency xxxx',
'agent',
'ago',
'agree',
'agreement',
'alert',
'allege',
'allege debt',
'allow',
'america',
'american',

'annual',
'answer',
'anymore',
'apartment',
'apparently',
'appear',
'application',
'apply',
'apply loan',
'appreciate',
'approval',
'approve',
'approximately',
'arrive',
'asap',
'ask',
'ask speak',
'ask xxxx',
'assign',
'assist',
'assistance',
'associate',
'assume',
'assure',
'attach',
'attempt',
'attempt collect',
'attention',
'attorney',
'authorization',
'authorize',
'auto',
'automate',
'automatic',
'automatically',
'avail',
'available',
'avoid',
'aware',
'away',
'bad',
'balance',
'balance xxxx',
'bank',
'bank account',
'bank america',
'bank xxxx',
'bankruptcy',
'base',

'basically',
'basis',
'bear',
'begin',
'behalf',
'believe',
'belong',
'benefit',
'best',
'better',
'block',
'branch',
'breach',
'break',
'bring',
'bureau',
'bureaus',
'bureaus xxxx',
'business',
'business days',
'buy',
'california',
'cancel',
'capital',
'car',
'card',
'card account',
'card company',
'card xxxx',
'care',
'case',
'case number',
'case xxxx',
'cash',
'catch',
'cause',
'cease',
'cell',
'cell phone',
'center',
'certain',
'certify',
'certify letter',
'certify mail',
'cfpb',
'change',
'chapter',
'chapter xxxx',
'charge',

'charge account',
'charge fee',
'charge xxxx',
'chase',
'chase bank',
'check',
'check account',
'check credit',
'check xxxx',
'choose',
'citi',
'citibank',
'civil',
'claim',
'clear',
'clearly',
'client',
'close',
'close account',
'close xxxx',
'code',
'collect',
'collect debt',
'collection',
'collection account',
'collection agency',
'collection company',
'collection practice',
'collections',
'collector',
'college',
'come',
'commit',
'communicate',
'communication',
'company',
'company xxxx',
'complain',
'complaint',
'complaint xxxx',
'complaints',
'complete',
'completely',
'compliance',
'comply',
'computer',
'concern',
'condition',
'conduct',

'confirm',
'confirmation',
'confuse',
'consent',
'consider',
'consumer',
'consumer financial',
'consumers',
'contact',
'contact company',
'contact xxxx',
'contain',
'continue',
'continue report',
'contract',
'control',
'conversation',
'copy',
'copy credit',
'correct',
'correctly',
'correspondence',
'cost',
'county',
'couple',
'course',
'court',
'cover',
'create',
'credit account',
'credit agencies',
'credit bureau',
'credit bureaus',
'credit card',
'credit file',
'credit history',
'credit report',
'credit score',
'credit xxxx',
'creditor',
'creditor xxxx',
'creditors',
'current',
'currently',
'customer',
'customer service',
'customers',
'daily',
'damage',

'damage credit',
'data',
'date',
'date xxxx',
'day',
'day xxxx',
'days',
'days late',
'days later',
'deal',
'dear',
'debit',
'debit card',
'debt',
'debt collection',
'debt collector',
'debt credit',
'debt owe',
'debt pay',
'debt xxxx',
'debts',
'deceptive',
'decide',
'decision',
'decline',
'default',
'delay',
'delete',
'delete credit',
'delinquency',
'delinquent',
'deliver',
'demand',
'deny',
'department',
'deposit',
'dept',
'derogatory',
'despite',
'determine',
'didn't',
'different',
'difficult',
'direct',
'directly',
'discharge',
'disclose',
'discover',
'discuss',

'dispute',
'dispute account',
'dispute credit',
'dispute xxxx',
'document',
'documentation',
'dollars',
'dont',
'double',
'drive',
'drop',
'earlier',
'early',
'education',
'effect',
'effort',
'eligible',
'email',
'email xxxx',
'employee',
'employees',
'employer',
'employment',
'end',
'end xxxx',
'ensure',
'enter',
'entire',
'equifax',
'equifax xxxx',
'equity',
'erroneous',
'error',
'errors',
'escalate',
'escrow',
'especially',
'establish',
'evidence',
'exact',
'exist',
'expect',
'expense',
'experian',
'experian xxxx',
'experience',
'expire',
'explain',
'explanation',

'express',
'extend',
'extra',
'extremely',
'fact',
'fail',
'failure',
'fair',
'fair credit',
'fair debt',
'faith',
'fall',
'false',
'family',
'far',
'fargo',
'fault',
'fax',
'fcra',
'fdcpa',
'federal',
'federal law',
'fee',
'fee xxxx',
'feel',
'felt',
'figure',
'file',
'file bankruptcy',
'file complaint',
'file dispute',
'file xxxx',
'final',
'finally',
'finance',
'financial',
'fine',
'firm',
'fix',
'follow',
'forbearance',
'force',
'foreclose',
'foreclosure',
'form',
'formal',
'forth',
'forward',
'fraud',

'fraud alert',
'fraudulent',
'fraudulent account',
'fraudulently',
'free',
'freeze',
'friday',
'frustrate',
'ftc',
'fully',
'fund',
'furthermore',
'future',
'general',
'good',
'government',
'grant',
'great',
'half',
'hand',
'handle',
'hang',
'happen',
'harass',
'harassment',
'hard',
'hardship',
'health',
'hear',
'hello',
'help',
'high',
'higher',
'hire',
'history',
'hit',
'hold',
'holder',
'home',
'home loan',
'home xxxx',
'honor',
'hop',
'hope',
'hospital',
'hour',
'hours',
'house',
'hurt',

'husband',
'idea',
'identify',
'identity',
'identity theft',
'ignore',
'illegal',
'illegally',
'immediately',
'impact',
'impossible',
'inaccurate',
'inaccurate information',
'include',
'income',
'incorrect',
'incorrect information',
'incorrectly',
'increase',
'incur',
'indicate',
'individual',
'info',
'inform',
'information',
'information credit',
'information provide',
'information report',
'information xxx',
'initial',
'initially',
'initiate',
'inquire',
'inquiries',
'inquiry',
'insist',
'instead',
'institution',
'instruct',
'insurance',
'insurance company',
'investigate',
'investigation',
'involve',
'issue',
'issue xxx',
'item',
'items',
'ive',
'job',

'kind',
'know',
'knowledge',
'lack',
'lady',
'large',
'late',
'late fee',
'late payment',
'late payments',
'late xxxx',
'later',
'law',
'laws',
'lawsuit',
'lawyer',
'lead',
'learn',
'lease',
'leave',
'leave message',
'legal',
'legal action',
'legally',
'lend',
'lender',
'let',
'let know',
'letter',
'letter date',
'letter say',
'letter send',
'letter state',
'letter xxxx',
'license',
'lie',
'lien',
'life',
'like',
'limit',
'line',
'link',
'list',
'list credit',
'little',
'live',
'live xxxx',
'llc',
'loan',

'loan modification',
'loan pay',
'loan service',
'loan xxxx',
'local',
'locate',
'lock',
'log',
'long',
'longer',
'look',
'lose',
'loss',
'lot',
'low',
'lower',
'mail',
'mail xxxx',
'maintain',
'major',
'make',
'make payment',
'make payments',
'make sure',
'make xxxx',
'management',
'manager',
'manager xxxx',
'manner',
'mark',
'market',
'match',
'matter',
'mean',
'medical',
'meet',
'member',
'mention',
'message',
'minimum',
'minutes',
'mislead',
'miss',
'miss payment',
'mistake',
'modification',
'modify',
'monday',
'money',

'money account',
'monitor',
'month',
'month xxxx',
'monthly',
'monthly payment',
'monthly payments',
'months',
'months ago',
'months xxxx',
'morning',
'mortgage',
'mortgage company',
'mortgage loan',
'mortgage payment',
'mortgage payments',
'mortgage xxxx',
'mother',
'multiple',
'multiple time',
'national',
'navient',
'nearly',
'necessary',
'need',
'need help',
'need remove',
'negative',
'negatively',
'new',
'non',
'note',
'notice',
'notification',
'notify',
'number',
'number xxxx',
'numerous',
'numerous time',
'obligation',
'obtain',
'obviously',
'occasion',
'occur',
'offer',
'office',
'officer',
'old',
'online',

'open',
'open account',
'open xxxx',
'opportunity',
'option',
'options',
'order',
'original',
'original creditor',
'originally',
'outstanding',
'overdraft',
'owe',
'owe debt',
'owe money',
'owe xxxx',
'owner',
'page',
'paper',
'paperwork',
'party',
'pass',
'past',
'pay',
'pay account',
'pay balance',
'pay debt',
'pay loan',
'pay month',
'pay mortgage',
'pay time',
'pay xxxx',
'payment',
'payment history',
'payment make',
'payment plan',
'payment xxxx',
'payments',
'payments make',
'payments time',
'payments xxxx',
'payoff',
'pending',
'people',
'period',
'permission',
'person',
'personal',
'personal information',

'phone',
'phone number',
'phone xxxx',
'pick',
'place',
'plan',
'plus',
'point',
'police',
'police report',
'policy',
'portfolio',
'possible',
'post',
'practice',
'practice act',
'present',
'prevent',
'previous',
'previously',
'price',
'principal',
'print',
'prior',
'private',
'problem',
'problems',
'proceed',
'process',
'profile',
'program',
'promise',
'promptly',
'proof',
'proper',
'properly',
'property',
'protect',
'protection',
'prove',
'provide',
'provide copy',
'provide information',
'provide proof',
'public',
'pull',
'pull credit',
'purchase',
'purchase home',

'purchase xxxx',
'purpose',
'pursuant',
'pursue',
'qualify',
'question',
'raise',
'rat',
'rate',
'reach',
'read',
'real',
'realize',
'really',
'reason',
'reasonable',
'receipt',
'receive',
'receive email',
'receive letter',
'receive notice',
'receive phone',
'receive response',
'receive xxxx',
'recent',
'recently',
'recognize',
'record',
'recovery',
'reduce',
'refer',
'reference',
'refinance',
'reflect',
'refund',
'refuse',
'refuse remove',
'regard',
'register',
'regular',
'relate',
'release',
'remain',
'removal',
'remove',
'remove account',
'remove credit',
'remove xxxx',
'rent',

'rep',
'repair',
'repayment',
'repeat',
'repeatedly',
'reply',
'report account',
'report act',
'report agencies',
'report agency',
'report company',
'report credit',
'report information',
'report late',
'report report',
'report xxxx',
'represent',
'representative',
'representative xxxx',
'representatives',
'request',
'request information',
'request xxxx',
'require',
'requirements',
'research',
'resolution',
'resolve',
'resolve issue',
'respond',
'response',
'responsibility',
'responsible',
'result',
'return',
'reverse',
'review',
'right',
'risk',
'rude',
'ruin',
'rule',
'run',
'sale',
'satisfy',
'save',
'saw',
'say',
'say owe',

'say xxxx',
'scam',
'schedule',
'school',
'score',
'score xxxx',
'second',
'section',
'secure',
'security',
'security number',
'seek',
'sell',
'send',
'send certify',
'send copy',
'send email',
'send letter',
'send xxxx',
'sense',
'separate',
'serve',
'service',
'service xxxx',
'servicer',
'set',
'settle',
'settlement',
'share',
'short',
'sign',
'signature',
'similar',
'simply',
'sincerely',
'single',
'sit',
'site',
'situation',
'small',
'social',
'social security',
'son',
'soon',
'speak',
'speak supervisor',
'speak xxxx',
'specific',
'specifically',

'spend',
'stand',
'start',
'start xxxx',
'state',
'state account',
'state owe',
'state xxxx',
'statement',
'statements',
'status',
'stay',
'steal',
'step',
'stop',
'store',
'stress',
'student',
'student loan',
'subject',
'submit',
'sue',
'suffer',
'suggest',
'suit',
'supervisor',
'supply',
'support',
'suppose',
'sure',
'switch',
'systems',
'talk',
'tax',
'telephone',
'tell',
'tell account',
'tell need',
'tell pay',
'tell xxxx',
'term',
'texas',
'thank',
'thank xxxx',
'theft',
'thing',
'things',
'think',
'threaten',

'time',
'time day',
'time tell',
'time xxxx',
'timely',
'title',
'today',
'today xxxx',
'total',
'track',
'trade',
'transaction',
'transactions',
'transfer',
'transfer xxxx',
'transunion',
'transunion xxxx',
'treat',
'true',
'trust',
'try',
'try collect',
'try contact',
'turn',
'twice',
'type',
'unable',
'unacceptable',
'unauthorized',
'understand',
'unfair',
'unfortunately',
'union',
'unknown',
'unless',
'unpaid',
'update',
'upset',
'use',
'use card',
'use xxxx',
'valid',
'validate',
'validate debt',
'validation',
'value',
'various',
'vehicle',
'verification',
'verify',

```

'verify account',
'verify information',
'victim',
'victim identity',
'view',
'violate',
'violation',
...]

```

```
len(vect.get_feature_names())
```

```
1106
```

```
x_train_tf=vect.transform(train_x)
```

```
x_train_tf.toarray()
```

```

array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)

```

```

x_train=pd.DataFrame(x_train_tf.toarray(),
columns=vect.get_feature_names())

```

```
x_train
```

	ability	able	absolutely	accept	access	accord	account
account \							
0	0	0	0	0	0	0	
0							
1	0	0	0	0	0	0	
0							
2	0	0	0	0	0	0	
0							
3	0	0	0	0	0	0	
0							
4	0	0	0	0	0	0	
0							
...	
...							
95886	0	0	0	0	0	0	2
0							
95887	0	0	0	0	0	0	0
0							
95888	0	0	0	0	0	0	0
0							
95889	0	0	0	0	0	0	0

```

0
95890      0      0      0      1      0      0
0
      account balance  account bank  account charge  ...  xxxxxxxx  \
0              0              0              0  ...      0
1              0              0              0  ...      0
2              0              0              0  ...      0
3              0              0              0  ...      0
4              0              0              0  ...      0
...          ...          ...          ...  ...      ...
95886          0              0              0  ...      0
95887          0              0              0  ...      0
95888          0              0              0  ...      0
95889          0              0              0  ...      0
95890          0              0              0  ...      0

```

```

      xxxxxxxx  xxxx  year  year  xxxx  years  years ago  years old  \
0              0      0      0      0      0      0      0
1              0      0      0      0      0      0      0
2              0      0      0      0      0      0      0
3              0      0      0      0      0      0      0
4              0      0      0      1      0      0      0
...          ...      ...      ...      ...      ...      ...
95886          0      0      0      0      0      0      0
95887          0      0      0      0      0      0      0
95888          0      0      0      0      0      0      0
95889          0      0      0      1      0      0      0
95890          0      0      0      0      0      0      0

```

```

      years  xxxx  yes  yesterday
0              0      0      0
1              0      0      0
2              0      0      0
3              0      0      0
4              0      0      0
...          ...      ...      ...
95886          0      0      0
95887          0      0      0
95888          0      0      0
95889          0      0      0
95890          0      0      0

```

```
[95891 rows x 1106 columns]
```

```

x_test_tf=vect.transform(val_x)
x_test=pd.DataFrame(x_test_tf.toarray(),
columns=vect.get_feature_names())

```

```
train_y.value_counts()
```

```

Credit_Reporting      31078
Debt_collection       21733
Mortgage              13231
Loan                  10778
cards                 10368
Bank_Accounts         6977
Money_Service_or_Currency_Service  1726
Name: Product, dtype: int64

y_train=train_y.map({"Credit_Reporting":0,"Debt_collection":1 ,"Mortgage":2, "Loan":3 , "cards":4,"Bank_Accounts":5,
"Money_Service_or_Currency_Service":6})
y_test=val_y.map({"Credit_Reporting":0,"Debt_collection":1 ,"Mortgage":2, "Loan":3 , "cards":4,"Bank_Accounts":5,
"Money_Service_or_Currency_Service":6})

y_train.value_counts()

0      31078
1      21733
2      13231
3      10778
4      10368
5       6977
6       1726
Name: Product, dtype: int64

from sklearn.tree import DecisionTreeClassifier
dt=DecisionTreeClassifier(max_depth=10)
dt.fit(x_train, y_train)

DecisionTreeClassifier(max_depth=10)

print("trainScore", dt.score(x_train, y_train))
print("testScore", dt.score(x_test, y_test))

trainScore 0.7094304992126477
testScore 0.6952268721777852

```

TF-IDF

```

from sklearn.feature_extraction.text import TfidfVectorizer
train_x, val_x, train_y, val_y= train_test_split(X,y ,test_size = 0.5,
random_state = 42)

y_train=train_y.map({"Credit_Reporting":0,"Debt_collection":1 ,"Mortgage":2, "Loan":3 , "cards":4,"Bank_Accounts":5,
"Money_Service_or_Currency_Service":6})
y_test=val_y.map({"Credit_Reporting":0,"Debt_collection":1 ,"Mortgage":2, "Loan":3 , "cards":4,"Bank_Accounts":5,
"Money_Service_or_Currency_Service":6})

```

```
tf=TfidfVectorizer(lowercase=True,stop_words="english",
                    ngram_range=(1, 2),max_df=.4,min_df=1000)

text_tf= tf.fit(train_x)

text_tf.get_feature_names()

['ability',
 'able',
 'absolutely',
 'accept',
 'access',
 'accord',
 'account account',
 'account balance',
 'account bank',
 'account charge',
 'account close',
 'account credit',
 'account information',
 'account list',
 'account number',
 'account open',
 'account pay',
 'account remove',
 'account report',
 'account tell',
 'account xxxx',
 'accrue',
 'acct',
 'accuracy',
 'accurate',
 'acknowledge',
 'act',
 'action',
 'active',
 'activity',
 'actual',
 'actually',
 'add',
 'addition',
 'additional',
 'additionally',
 'address',
 'address xxxx',
 'admit',
 'advance',
 'advantage',
 'advise',
 'affect',
```

'affect credit',
'affidavit',
'afford',
'age',
'agencies',
'agency',
'agency xxxx',
'agent',
'ago',
'agree',
'agreement',
'alert',
'allege',
'allege debt',
'allow',
'america',
'american',
'annual',
'answer',
'anymore',
'apartment',
'apparently',
'appear',
'application',
'apply',
'apply loan',
'appreciate',
'approval',
'approve',
'approximately',
'arrive',
'asap',
'ask',
'ask speak',
'ask xxxx',
'assign',
'assist',
'assistance',
'associate',
'assume',
'assure',
'attach',
'attempt',
'attempt collect',
'attention',
'attorney',
'authorization',
'authorize',
'auto',

'automate',
'automatic',
'automatically',
'avail',
'available',
'avoid',
'aware',
'away',
'bad',
'balance',
'balance xxxx',
'bank',
'bank account',
'bank america',
'bank xxxx',
'bankruptcy',
'base',
'basically',
'basis',
'bear',
'begin',
'behalf',
'believe',
'belong',
'benefit',
'best',
'better',
'block',
'branch',
'breach',
'break',
'bring',
'bureau',
'bureaus',
'bureaus xxxx',
'business',
'business days',
'buy',
'california',
'cancel',
'capital',
'car',
'card',
'card account',
'card company',
'card xxxx',
'care',
'case',
'case number',

'case xxxx',
'cash',
'catch',
'cause',
'cease',
'cell',
'cell phone',
'center',
'certain',
'certify',
'certify letter',
'certify mail',
'cfpb',
'change',
'chapter',
'chapter xxxx',
'charge',
'charge account',
'charge fee',
'charge xxxx',
'chase',
'chase bank',
'check',
'check account',
'check credit',
'check xxxx',
'choose',
'citi',
'citibank',
'civil',
'claim',
'clear',
'clearly',
'client',
'close',
'close account',
'close xxxx',
'code',
'collect',
'collect debt',
'collection',
'collection account',
'collection agency',
'collection company',
'collection practice',
'collections',
'collector',
'college',
'come',

'commit',
'communicate',
'communication',
'company',
'company xxxx',
'complain',
'complaint',
'complaint xxxx',
'complaints',
'complete',
'completely',
'compliance',
'comply',
'computer',
'concern',
'condition',
'conduct',
'confirm',
'confirmation',
'confuse',
'consent',
'consider',
'consumer',
'consumer financial',
'consumers',
'contact',
'contact company',
'contact xxxx',
'contain',
'continue',
'continue report',
'contract',
'control',
'conversation',
'copy',
'copy credit',
'correct',
'correctly',
'correspondence',
'cost',
'county',
'couple',
'course',
'court',
'cover',
'create',
'credit account',
'credit agencies',
'credit bureau',

'credit bureaus',
'credit card',
'credit file',
'credit history',
'credit report',
'credit score',
'credit xxxx',
'creditor',
'creditor xxxx',
'creditors',
'current',
'currently',
'customer',
'customer service',
'customers',
'daily',
'damage',
'damage credit',
'data',
'date',
'date xxxx',
'day',
'day xxxx',
'days',
'days late',
'days later',
'deal',
'dear',
'debit',
'debit card',
'debt',
'debt collection',
'debt collector',
'debt credit',
'debt owe',
'debt pay',
'debt xxxx',
'debts',
'deceptive',
'decide',
'decision',
'decline',
'default',
'delay',
'delete',
'delete credit',
'delinquency',
'delinquent',
'deliver',

'demand',
'deny',
'department',
'deposit',
'dept',
'derogatory',
'despite',
'determine',
'didn't',
'different',
'difficult',
'direct',
'directly',
'discharge',
'disclose',
'discover',
'discuss',
'dispute',
'dispute account',
'dispute credit',
'dispute xxxx',
'document',
'documentation',
'dollars',
'don't',
'double',
'drive',
'drop',
'earlier',
'early',
'education',
'effect',
'effort',
'eligible',
'email',
'email xxxx',
'employee',
'employees',
'employer',
'employment',
'end',
'end xxxx',
'ensure',
'enter',
'entire',
'equifax',
'equifax xxxx',
'equity',
'erroneous',

'error',
'errors',
'escalate',
'escrow',
'especially',
'establish',
'evidence',
'exact',
'exist',
'expect',
'expense',
'experian',
'experian xxxx',
'experience',
'expire',
'explain',
'explanation',
'express',
'extend',
'extra',
'extremely',
'fact',
'fail',
'failure',
'fair',
'fair credit',
'fair debt',
'faith',
'fall',
'false',
'family',
'far',
'fargo',
'fault',
'fax',
'fcra',
'fdcpa',
'federal',
'federal law',
'fee',
'fee xxxx',
'feel',
'felt',
'figure',
'file',
'file bankruptcy',
'file complaint',
'file dispute',
'file xxxx',

'final',
'finally',
'finance',
'financial',
'fine',
'firm',
'fix',
'follow',
'forbearance',
'force',
'foreclose',
'foreclosure',
'form',
'formal',
'forth',
'forward',
'fraud',
'fraud alert',
'fraudulent',
'fraudulent account',
'fraudulently',
'free',
'freeze',
'friday',
'frustrate',
'ftc',
'fully',
'fund',
'furthermore',
'future',
'general',
'good',
'government',
'grant',
'great',
'half',
'hand',
'handle',
'hang',
'happen',
'harass',
'harassment',
'hard',
'hardship',
'health',
'hear',
'hello',
'help',
'high',

'higher',
'hire',
'history',
'hit',
'hold',
'holder',
'home',
'home loan',
'home xxxx',
'honor',
'hop',
'hope',
'hospital',
'hour',
'hours',
'house',
'hurt',
'husband',
'idea',
'identify',
'identity',
'identity theft',
'ignore',
'illegal',
'illegally',
'immediately',
'impact',
'impossible',
'inaccurate',
'inaccurate information',
'include',
'income',
'incorrect',
'incorrect information',
'incorrectly',
'increase',
'incur',
'indicate',
'individual',
'info',
'inform',
'information',
'information credit',
'information provide',
'information report',
'information xxxx',
'initial',
'initially',
'initiate',

'inquire',
'inquiries',
'inquiry',
'insist',
'instead',
'institution',
'instruct',
'insurance',
'insurance company',
'investigate',
'investigation',
'involve',
'issue',
'issue xxxx',
'item',
'items',
'ive',
'job',
'kind',
'know',
'knowledge',
'lack',
'lady',
'large',
'late',
'late fee',
'late payment',
'late payments',
'late xxxx',
'later',
'law',
'laws',
'lawsuit',
'lawyer',
'lead',
'learn',
'lease',
'leave',
'leave message',
'legal',
'legal action',
'legally',
'lend',
'lender',
'let',
'let know',
'letter',
'letter date',
'letter say',
'letter send',

'letter state',
'letter xxxx',
'license',
'lie',
'lien',
'life',
'like',
'limit',
'line',
'link',
'list',
'list credit',
'little',
'live',
'live xxxx',
'llc',
'loan',
'loan modification',
'loan pay',
'loan service',
'loan xxxx',
'local',
'locate',
'lock',
'log',
'long',
'longer',
'look',
'lose',
'loss',
'lot',
'low',
'lower',
'mail',
'mail xxxx',
'maintain',
'major',
'make',
'make payment',
'make payments',
'make sure',
'make xxxx',
'management',
'manager',
'manager xxxx',
'manner',
'mark',
'market',
'match',

'matter',
'mean',
'medical',
'meet',
'member',
'mention',
'message',
'minimum',
'minutes',
'mislead',
'miss',
'miss payment',
'mistake',
'modification',
'modify',
'monday',
'money',
'money account',
'monitor',
'month',
'month xxxx',
'monthly',
'monthly payment',
'monthly payments',
'months',
'months ago',
'months xxxx',
'morning',
'mortgage',
'mortgage company',
'mortgage loan',
'mortgage payment',
'mortgage payments',
'mortgage xxxx',
'mother',
'multiple',
'multiple time',
'national',
'navient',
'nearly',
'necessary',
'need',
'need help',
'need remove',
'negative',
'negatively',
'new',
'non',
'note',

'notice',
'notification',
'notify',
'number',
'number xxxx',
'numerous',
'numerous time',
'obligation',
'obtain',
'obviously',
'occasion',
'occur',
'offer',
'office',
'officer',
'old',
'online',
'open',
'open account',
'open xxxx',
'opportunity',
'option',
'options',
'order',
'original',
'original creditor',
'originally',
'outstanding',
'overdraft',
'owe',
'owe debt',
'owe money',
'owe xxxx',
'owner',
'page',
'paper',
'paperwork',
'party',
'pass',
'past',
'pay',
'pay account',
'pay balance',
'pay debt',
'pay loan',
'pay month',
'pay mortgage',
'pay time',
'pay xxxx',

'payment',
'payment history',
'payment make',
'payment plan',
'payment xxxx',
'payments',
'payments make',
'payments time',
'payments xxxx',
'payoff',
'pending',
'people',
'period',
'permission',
'person',
'personal',
'personal information',
'phone',
'phone number',
'phone xxxx',
'pick',
'place',
'plan',
'plus',
'point',
'police',
'police report',
'policy',
'portfolio',
'possible',
'post',
'practice',
'practice act',
'present',
'prevent',
'previous',
'previously',
'price',
'principal',
'print',
'prior',
'private',
'problem',
'problems',
'proceed',
'process',
'profile',
'program',
'promise',

'promptly',
'proof',
'proper',
'properly',
'property',
'protect',
'protection',
'prove',
'provide',
'provide copy',
'provide information',
'provide proof',
'public',
'pull',
'pull credit',
'purchase',
'purchase home',
'purchase xxxx',
'purpose',
'pursuant',
'pursue',
'qualify',
'question',
'raise',
'rat',
'rate',
'reach',
'read',
'real',
'realize',
'really',
'reason',
'reasonable',
'receipt',
'receive',
'receive email',
'receive letter',
'receive notice',
'receive phone',
'receive response',
'receive xxxx',
'recent',
'recently',
'recognize',
'record',
'recovery',
'reduce',
'refer',
'reference',

'refinance',
'reflect',
'refund',
'refuse',
'refuse remove',
'regard',
'register',
'regular',
'relate',
'release',
'remain',
'removal',
'remove',
'remove account',
'remove credit',
'remove xxxx',
'rent',
'rep',
'repair',
'repayment',
'repeat',
'repeatedly',
'reply',
'report account',
'report act',
'report agencies',
'report agency',
'report company',
'report credit',
'report information',
'report late',
'report report',
'report xxxx',
'represent',
'representative',
'representative xxxx',
'representatives',
'request',
'request information',
'request xxxx',
'require',
'requirements',
'research',
'resolution',
'resolve',
'resolve issue',
'respond',
'response',
'responsibility',

'responsible',
'result',
'return',
'reverse',
'review',
'right',
'risk',
'rude',
'ruin',
'rule',
'run',
'sale',
'satisfy',
'save',
'saw',
'say',
'say owe',
'say xxxx',
'scam',
'schedule',
'school',
'score',
'score xxxx',
'second',
'section',
'secure',
'security',
'security number',
'seek',
'sell',
'send',
'send certify',
'send copy',
'send email',
'send letter',
'send xxxx',
'sense',
'separate',
'serve',
'service',
'service xxxx',
'servicer',
'set',
'settle',
'settlement',
'share',
'short',
'sign',
'signature',

'similar',
'simply',
'sincerely',
'single',
'sit',
'site',
'situation',
'small',
'social',
'social security',
'son',
'soon',
'speak',
'speak supervisor',
'speak xxxx',
'specific',
'specifically',
'spend',
'stand',
'start',
'start xxxx',
'state',
'state account',
'state owe',
'state xxxx',
'statement',
'statements',
'status',
'stay',
'steal',
'step',
'stop',
'store',
'stress',
'student',
'student loan',
'subject',
'submit',
'sue',
'suffer',
'suggest',
'suit',
'supervisor',
'supply',
'support',
'suppose',
'sure',
'switch',
'systems',

'talk',
'tax',
'telephone',
'tell',
'tell account',
'tell need',
'tell pay',
'tell xxxx',
'term',
'texas',
'thank',
'thank xxxx',
'theft',
'thing',
'things',
'think',
'threaten',
'time',
'time day',
'time tell',
'time xxxx',
'timely',
'title',
'today',
'today xxxx',
'total',
'track',
'trade',
'transaction',
'transactions',
'transfer',
'transfer xxxx',
'transunion',
'transunion xxxx',
'treat',
'true',
'trust',
'try',
'try collect',
'try contact',
'turn',
'twice',
'type',
'unable',
'unacceptable',
'unauthorized',
'understand',
'unfair',
'unfortunately',

xxxxxxx	xxxx	year	year	xxxx	years	years	ago	years	old	years
xxxx \										
0	0.0	0.0		0.0	0.0		0.0		0.0	

```

0.0
1      0.0   0.0      0.0   0.0      0.0      0.0
0.0

```

```

    yes yesterday
0  0.0      0.0
1  0.0      0.0

```

[2 rows x 1106 columns]

train_x_tf_df

	ability	able	absolutely	accept	access	accord	account
account \							
0	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
1	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
2	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
3	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
4	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
...
...							
95886	0.0	0.0	0.0	0.000000	0.0	0.27632	
0.0							
95887	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
95888	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
95889	0.0	0.0	0.0	0.000000	0.0	0.000000	
0.0							
95890	0.0	0.0	0.0	0.142097	0.0	0.000000	
0.0							

	account balance	account bank	account charge	...	xxxxxxxxx \
0	0.0	0.0	0.0	...	0.0
1	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	...	0.0
...
95886	0.0	0.0	0.0	...	0.0
95887	0.0	0.0	0.0	...	0.0
95888	0.0	0.0	0.0	...	0.0
95889	0.0	0.0	0.0	...	0.0
95890	0.0	0.0	0.0	...	0.0

	xxxxxxxx	xxxx	year	year	xxxx	years	years	ago	years	old
\										
0		0.0	0.0		0.0	0.000000		0.0		0.0
1		0.0	0.0		0.0	0.000000		0.0		0.0
2		0.0	0.0		0.0	0.000000		0.0		0.0
3		0.0	0.0		0.0	0.000000		0.0		0.0
4		0.0	0.0		0.0	0.059053		0.0		0.0
...	
95886		0.0	0.0		0.0	0.000000		0.0		0.0
95887		0.0	0.0		0.0	0.000000		0.0		0.0
95888		0.0	0.0		0.0	0.000000		0.0		0.0
95889		0.0	0.0		0.0	0.256971		0.0		0.0
95890		0.0	0.0		0.0	0.000000		0.0		0.0

	years	xxxx	yes	yesterday
0		0.0	0.0	0.0
1		0.0	0.0	0.0
2		0.0	0.0	0.0
3		0.0	0.0	0.0
4		0.0	0.0	0.0
...	
95886		0.0	0.0	0.0
95887		0.0	0.0	0.0
95888		0.0	0.0	0.0
95889		0.0	0.0	0.0
95890		0.0	0.0	0.0

[95891 rows x 1106 columns]

```
val_x_tf=text_tf.transform(val_x)
val_x_tf_df=pd.DataFrame(val_x_tf.toarray(),columns=text_tf.get_feature_names())
val_x_tf_df.head(2)
```

	ability	able	absolutely	accept	access	accord	account	account
\								
0	0.0	0.0	0.0	0.0	0.0	0.0		0.0
1	0.0	0.0	0.0	0.0	0.0	0.0		0.0

	account	balance	account	bank	account	charge	...	xxxxxxx	\
0		0.0		0.0		0.0	...	0.0	
1		0.0		0.0		0.0	...	0.0	

	xxxxxxx	xxxx	year	year	xxxx	years	years	ago	years	old	years
0		0.0	0.0		0.0	0.0		0.0		0.0	
1		0.0	0.0		0.0	0.0		0.0		0.0	

	yes	yesterday
0	0.0	0.0
1	0.0	0.0

[2 rows x 1106 columns]

```
dt_tf=DecisionTreeClassifier(max_depth=10)
dt_tf.fit(train_x_tf_df, y_train)

DecisionTreeClassifier(max_depth=10)

print("Train Score:", dt_tf.score(train_x_tf_df, y_train))
print("val Score:", dt_tf.score(val_x_tf_df, y_test))

Train Score: 0.7110052038251765
val Score: 0.6932454557779145

sentence = ["""your debt deaprtment is group of croocks ,
They never deal a person with any diginity"""]

new=text_tf.transform(sentence)
new1=pd.DataFrame(new.toarray(), columns=text_tf.get_feature_names())
ppt=dt_tf.predict(new1)

if ppt[0]==0:
    print("Credit_Reporting")
elif ppt[0]==1:
    print("Debt_collection")
elif ppt[0]==2:
    print("Mortgage")
elif ppt[0]==3:
    print("Loan")
elif ppt[0]==4:
    print("Cards")
elif ppt[0]==5:
    print("Bank_Accounts")
elif ppt[0]==6:
```

```
print("Money_Service_or_Currency_Service")
```

Debt_collection