Customer Complain_Kaiburr

September 21, 2023

```
[1]: """
Perform a Text Classification on consumer complaint dataset
(https://catalog.data.gov/dataset/consumer-complaint-database) into following␣
  ↪categories.

0 Credit reporting, repair, or other
1 Debt collection
2 Consumer Loan
3 Mortgage

Steps to be followed -
1. Explanatory Data Analysis and Feature Engineering
2. Text Pre-Processing
3. Selection of Multi Classification model
4. Comparison of model performance
5. Model Evaluation
6. Prediction

"""
```

```
[1]: '\nPerform a Text Classification on consumer complaint
     dataset\n(https://catalog.data.gov/dataset/consumer-complaint-database) into
     following categories.\n\n0 Credit reporting, repair, or other\n1 Debt
     collection\n2 Consumer Loan\n3 Mortgage\n\nSteps to be followed -\n1.
     Explanatory Data Analysis and Feature Engineering\n2. Text Pre-Processing\n3.
     Selection of Multi Classification model\n4. Comparison of model performance\n5.
     Model Evaluation\n6. Prediction\n\n'
```

1 IMPORTING LIBRARIES

```
[2]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import nltk
     import string
```

```python
from scipy.stats import randint

from io import StringIO

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_selection import chi2
from IPython.display import display
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import LinearSVC
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion_matrix
from sklearn import metrics
```

2 DATA EXPLORATION

```python
[3]: df = pd.read_csv('complaints.csv')
     df.head().T
```

C:\Users\sjaru\AppData\Local\Temp\ipykernel_1440\719507999.py:1: DtypeWarning:
Columns (16) have mixed types. Specify dtype option on import or set
low_memory=False.
  df = pd.read_csv('complaints.csv')

```
[3]:                                                                    0
     \
     Date received                                             2023-08-25
     Product                       Credit reporting or other personal consumer re…
     Sub-product                                          Credit reporting
     Issue                                         Improper use of your report
     Sub-issue                         Reporting company used your report improperly
     Consumer complaint narrative                                     NaN
     Company public response                                          NaN
     Company                                    SANTANDER HOLDINGS USA, INC.
     State                                                             FL
     ZIP code                                                       33972
     Tags                                                             NaN
     Consumer consent provided?                                       NaN
     Submitted via                                                    Web
     Date sent to company                                      2023-08-25
     Company response to consumer                       Closed with explanation
     Timely response?                                                 Yes
     Consumer disputed?                                               NaN
     Complaint ID                                                 7512160
```

\
```
Date received                              2023-08-25
Product                  Credit reporting or other personal consumer re…
Sub-product                                   Credit reporting
Issue                           Incorrect information on your report
Sub-issue                      Information belongs to someone else
Consumer complaint narrative                            NaN
Company public response                                 NaN
Company                              WELLS FARGO & COMPANY
State                                                    CA
ZIP code                                              92394
Tags                                                    NaN
Consumer consent provided?                              NaN
Submitted via                                           Web
Date sent to company                           2023-08-25
Company response to consumer                   In progress
Timely response?                                        Yes
Consumer disputed?                                      NaN
Complaint ID                                      7466203
```

\
```
Date received                              2023-08-23
Product                  Credit reporting, credit repair services, or o…
Sub-product                                   Credit reporting
Issue                                 Improper use of your report
Sub-issue              Reporting company used your report improperly
Consumer complaint narrative                            NaN
Company public response                                 NaN
Company                       Experian Information Solutions Inc.
State                                                    WA
ZIP code                                              98507
Tags                                                    NaN
Consumer consent provided?                              NaN
Submitted via                                           Web
Date sent to company                           2023-08-23
Company response to consumer                   In progress
Timely response?                                        Yes
Consumer disputed?                                      NaN
Complaint ID                                      7442793
```

\
```
Date received                              2023-08-23
Product                  Credit reporting, credit repair services, or o…
```

```
Sub-product                                               Credit reporting
Issue                                           Improper use of your report
Sub-issue                       Reporting company used your report improperly
Consumer complaint narrative                                           NaN
Company public response                                                NaN
Company                               Experian Information Solutions Inc.
State                                                                   OR
ZIP code                                                             97209
Tags                                                                   NaN
Consumer consent provided?                                             NaN
Submitted via                                                          Web
Date sent to company                                            2023-08-23
Company response to consumer                                   In progress
Timely response?                                                       Yes
Consumer disputed?                                                     NaN
Complaint ID                                                       7442804

                                                                         4
Date received                                                   2023-08-23
Product                                       Credit card or prepaid card
Sub-product                     General-purpose credit card or charge card
Issue                                 Other features, terms, or problems
Sub-issue                                                   Other problem
Consumer complaint narrative                                           NaN
Company public response                                                NaN
Company                                                    CITIBANK, N.A.
State                                                                   WI
ZIP code                                                             53058
Tags                                                                   NaN
Consumer consent provided?                                             NaN
Submitted via                                                          Web
Date sent to company                                            2023-08-23
Company response to consumer                                   In progress
Timely response?                                                       Yes
Consumer disputed?                                                     NaN
Complaint ID                                                       7448208
```

[4]: `df.count()`

```
[4]: Date received                  4042985
     Product                        4042985
     Sub-product                    3807694
     Issue                          4042985
     Sub-issue                      3327227
     Consumer complaint narrative   1464248
     Company public response        1868451
     Company                        4042985
```

```
State                       3999820
ZIP code                    4012777
Tags                         423775
Consumer consent provided?  3142965
Submitted via               4042985
Date sent to company        4042985
Company response to consumer  4042980
Timely response?            4042985
Consumer disputed?           768358
Complaint ID                4042985
dtype: int64
```

[5]: `df.shape`

[5]: (4042985, 18)

[6]:
```python
# Create a new dataframe with two columns
df1 = df[['Product', 'Consumer complaint narrative']].copy()

# Remove missing values (NaN)
df1 = df1[pd.notnull(df1['Consumer complaint narrative'])]

# Renaming second column for a simpler name
df1.columns = ['Product', 'Consumer_complaint']

df1.shape
```

[6]: (1464248, 2)

[7]:
```python
# Percentage of complaints with text
total = df1['Consumer_complaint'].notnull().sum()
round((total/len(df)*100),1)
```

[7]: 36.2

[8]: `pd.DataFrame(df.Product.unique()).values`

[8]:
```
array([['Credit reporting or other personal consumer reports'],
       ['Credit reporting, credit repair services, or other personal consumer
reports'],
       ['Credit card or prepaid card'],
       ['Debt collection'],
       ['Mortgage'],
       ['Checking or savings account'],
       ['Vehicle loan or lease'],
       ['Credit card'],
       ['Payday loan, title loan, or personal loan'],
       ['Student loan'],
```

```
       ['Money transfer, virtual currency, or money service'],
       ['Payday loan, title loan, personal loan, or advance loan'],
       ['Debt or credit management'],
       ['Prepaid card'],
       ['Credit reporting'],
       ['Consumer Loan'],
       ['Bank account or service'],
       ['Payday loan'],
       ['Money transfers'],
       ['Other financial service'],
       ['Virtual currency']], dtype=object)
```
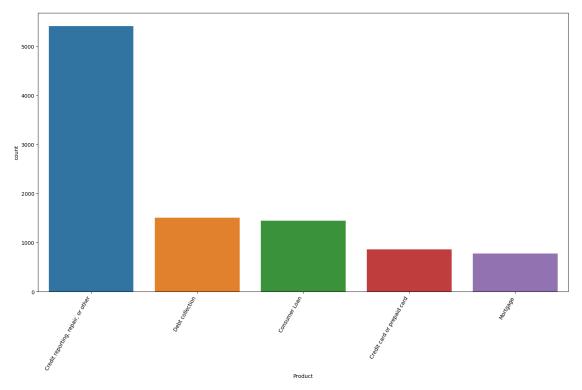
3 DATA PREPROCESSING

```python
[9]:  # Because the computation is time consuming (in terms of CPU), the data was␣
      ↪sampled
      df2 = df1.sample(10000, random_state=1).copy()
```

```python
[10]: # Renaming categories
      df2.replace({'Product':
                  {'Credit reporting or other personal consumer reports': 'Credit␣
      ↪reporting, repair, or other',
                   'Credit reporting, credit repair services, or other personal␣
      ↪consumer reports': 'Credit reporting, repair, or other',
                   'Credit card or prepaid card': 'Credit card or prepaid card',
                   'Debt collection': 'Debt collection',
                   'Mortgage': 'Mortgage',
                   'Checking or savings account': 'Consumer Loan',
                   'Vehicle loan or lease': 'Consumer Loan',
                   'Credit card': 'Credit card or prepaid card',
                   'Payday loan, title loan, or personal loan': 'Consumer Loan',
                   'Student loan': 'Consumer Loan',
                   'Money transfer, virtual currency, or money service': 'Consumer␣
      ↪Loan',
                   'Payday loan, title loan, personal loan, or advance loan':␣
      ↪'Consumer Loan',
                   'Debt or credit management': 'Credit reporting, repair, or␣
      ↪other',
                   'Prepaid card': 'Credit card or prepaid card',
                   'Credit reporting': 'Credit reporting, repair, or other',
                   'Consumer Loan': 'Consumer Loan',
                   'Bank account or service': 'Consumer Loan',
                   'Payday loan': 'Consumer Loan',
                   'Money transfers': 'Consumer Loan',
                   'Other financial service': 'Credit reporting, repair, or other',
                   'Virtual currency': 'Credit reporting, repair, or other'}},
```

```
                inplace= True)
```

```
[11]: pd.DataFrame(df2.Product.unique())
```

```
[11]:                                   0
      0  Credit reporting, repair, or other
      1                       Consumer Loan
      2                            Mortgage
      3             Credit card or prepaid card
      4                     Debt collection
```

```
[12]: # The number of classes were reduced from 22 to 5.
      # Now we need to represent each class as a number, so as our predictive model␣
       ↪can better understand the different categories.
```

4 VISUALISING PROCESSED DATA

```
[13]: # Create a new column 'category_id' with encoded categories
      df2['category_id'] = df2['Product'].factorize()[0]
      category_id_df = df2[['Product', 'category_id']].drop_duplicates()

      # Dictionaries for future use
      category_to_id = dict(category_id_df.values)
      id_to_category = dict(category_id_df[['category_id', 'Product']].values)

      # New dataframe
      df2.head()
```

```
[13]:                                Product  \
      2054782  Credit reporting, repair, or other
      1540532                       Consumer Loan
      2721658                            Mortgage
      1596292             Credit card or prepaid card
      580828                       Debt collection

                                  Consumer_complaint  category_id
      2054782  accordance with the Fair Credit Reporting act …            0
      1540532  I had an account since XXXX 2012 with capital …            1
      2721658  My name is XXXX XXXX XXXX, I am an XXXX  who r…            2
      1596292  Bank of America is denying credit based on fed…            3
      580828   I not only dispute the debt validity because t…            4
```

```
[14]: value_counts = df2['Product'].value_counts().sort_values(ascending=False)
      fig, ax = plt.subplots(figsize=(15, 10))
      sns.countplot(x='Product', data=df2, order=value_counts.index, ax=ax)
      ax.set_xticklabels(ax.get_xticklabels(), rotation=60, ha='right')
```

```
plt.tight_layout()
plt.show()

print(value_counts)
```
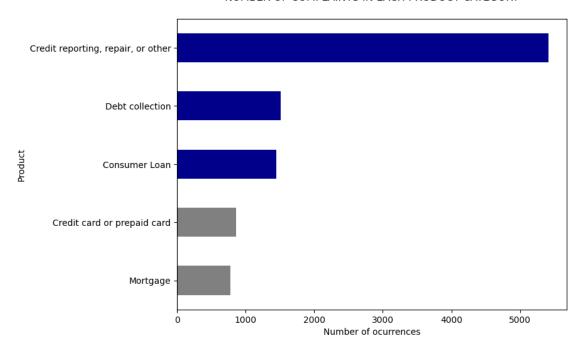


```
Product
Credit reporting, repair, or other      5410
Debt collection                         1509
Consumer Loan                           1447
Credit card or prepaid card              860
Mortgage                                 774
Name: count, dtype: int64
```

```
[15]: fig = plt.figure(figsize=(8,6))

colors = ['grey','grey','darkblue','darkblue','darkblue']

df2.groupby('Product').Consumer_complaint.count().sort_values().plot.barh(
    ylim=0, color=colors, title= 'NUMBER OF COMPLAINTS IN EACH PRODUCT␣
 ↪CATEGORY\n')

plt.xlabel('Number of ocurrences', fontsize = 10);
```

NUMBER OF COMPLAINTS IN EACH PRODUCT CATEGORY

5 TEXT DATA PREPROCESSING AND FEATURE ENGINEERING USING TF-IDF VECTORIZATION

```python
[16]: tfidf = TfidfVectorizer(sublinear_tf=True, min_df=5,
                              ngram_range=(1, 2),
                              stop_words='english')

      # We transform each complaint into a vector
      features = tfidf.fit_transform(df2.Consumer_complaint).toarray()

      labels = df2.category_id

      print("Each of the %d complaints is represented by %d features (TF-IDF score of␣
       ↪unigrams and bigrams)" %(features.shape))
```

Each of the 10000 complaints is represented by 26190 features (TF-IDF score of
unigrams and bigrams)

```python
[17]: # Number of top correlated terms to display
      N = 3

      # Loop through categories
      for Product, category_id in sorted(category_to_id.items()):
          features_chi2 = chi2(features, labels == category_id)
```

```python
    indices = np.argsort(features_chi2[0])

    feature_names = np.array(tfidf.get_feature_names_out())[indices]
    unigrams = [v for v in feature_names if len(v.split(' ')) == 1]
    bigrams = [v for v in feature_names if len(v.split(' ')) == 2]

    print("\n==> %s:" % (Product))
    print("  * Most Correlated Unigrams are: %s" % (', '.join(unigrams[-N:])))
    print("  * Most Correlated Bigrams are: %s" % (', '.join(bigrams[-N:])))
```

```
==> Consumer Loan:
  * Most Correlated Unigrams are: money, funds, bank
  * Most Correlated Bigrams are: wells fargo, overdraft fees, checking account

==> Credit card or prepaid card:
  * Most Correlated Unigrams are: merchant, citi, card
  * Most Correlated Bigrams are: card company, american express, credit card

==> Credit reporting, repair, or other:
  * Most Correlated Unigrams are: 1681, section, reporting
  * Most Correlated Bigrams are: credit report, 15 1681, 1681 section

==> Debt collection:
  * Most Correlated Unigrams are: collect, collection, debt
  * Most Correlated Bigrams are: collect debt, portfolio recovery, collection
agency

==> Mortgage:
  * Most Correlated Unigrams are: modification, escrow, mortgage
  * Most Correlated Bigrams are: escrow account, mortgage company, loan
modification
```

```python
[18]: X = df2['Consumer_complaint'] # Collection of documents
      y = df2['Product'] # Target or the labels we want to predict (i.e., the 5
       ↪different complaints of products)

      X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                          test_size=0.25,
                                                          random_state = 0)
```
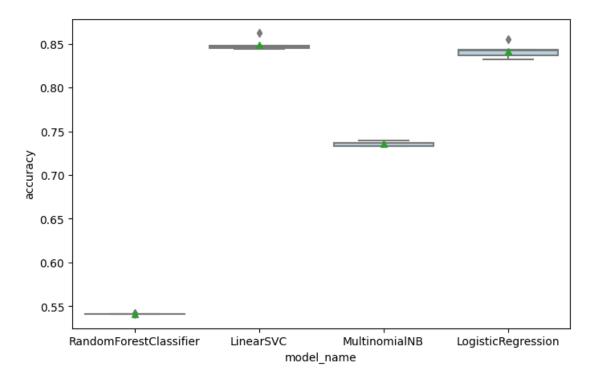
6 MODEL BUILDING AND EVALUATION

```python
[36]: models = [
          RandomForestClassifier(n_estimators=100, max_depth=5, random_state=0),
          LinearSVC(),
          MultinomialNB(),
```

```
        LogisticRegression(random_state=0, max_iter=1000)
]

# 5 Cross-validation
CV = 5
cv_df = pd.DataFrame(index=range(CV * len(models)))

entries = []
for model in models:
  model_name = model.__class__.__name__
  accuracies = cross_val_score(model, features, labels, scoring='accuracy',␣
  ↪cv=CV)

  for fold_idx, accuracy in enumerate(accuracies):
    entries.append((model_name, fold_idx, accuracy))

cv_df = pd.DataFrame(entries, columns=['model_name', 'fold_idx', 'accuracy'])
```

```
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
```

```
[37]: mean_accuracy = cv_df.groupby('model_name').accuracy.mean()
      std_accuracy = cv_df.groupby('model_name').accuracy.std()

      acc = pd.concat([mean_accuracy, std_accuracy], axis= 1,ignore_index=True)
      acc.columns = ['Mean Accuracy', 'Standard deviation']

      acc
```

```
[37]:                      Mean Accuracy   Standard deviation
      model_name
      LinearSVC                   0.8492             0.007555
      LogisticRegression          0.8419             0.008685
      MultinomialNB               0.7357             0.002752
      RandomForestClassifier      0.5415             0.000354
```

```python
[38]: plt.figure(figsize=(8,5))

      sns.boxplot(x='model_name', y='accuracy',
                  data=cv_df,
                  color='lightblue',
                  showmeans=True)

      plt.title("MEAN ACCURACY (cv = 5)\n", size=14);
```



```python
[ ]:
```

```python
[22]: from sklearn.model_selection import train_test_split

      X_train, X_test, y_train, y_test, indices_train, indices_test =␣
       ↪train_test_split(
```

```
        features, labels, df2.index, test_size=0.25,random_state=1, stratify=labels
)

model = LinearSVC()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(
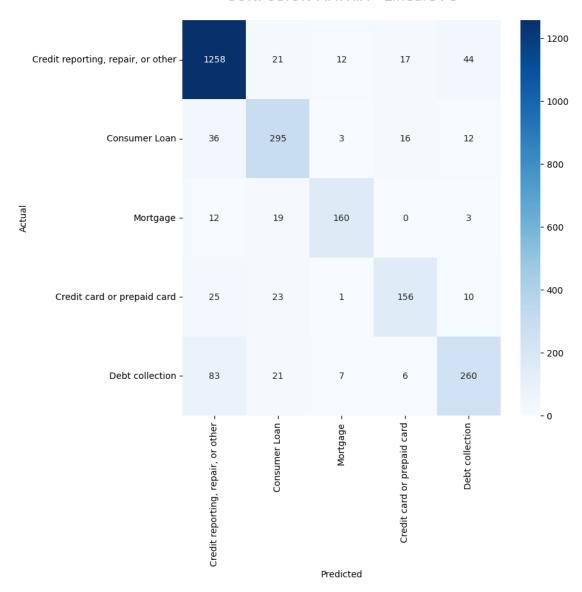
[23]:
```
# Classification report
print('\t\t\t\tCLASSIFICATIION METRICS\n')
print(metrics.classification_report(y_test, y_pred,target_names= df2['Product'].
  ↪unique()))
```

CLASSIFICATIION METRICS

	precision	recall	f1-score	support
Credit reporting, repair, or other	0.89	0.93	0.91	1352
Consumer Loan	0.78	0.81	0.80	362
Mortgage	0.87	0.82	0.85	194
Credit card or prepaid card	0.80	0.73	0.76	215
Debt collection	0.79	0.69	0.74	377
accuracy			0.85	2500
macro avg	0.83	0.80	0.81	2500
weighted avg	0.85	0.85	0.85	2500

[24]:
```
conf_mat = confusion_matrix(y_test, y_pred)
fig, ax = plt.subplots(figsize=(8,8))
sns.heatmap(conf_mat, annot=True, cmap="Blues", fmt='d',
            xticklabels=category_id_df.Product.values,
            yticklabels=category_id_df.Product.values)
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.title("CONFUSION MATRIX - LinearSVC\n", size=16);
```

CONFUSION MATRIX - LinearSVC


```
[25]: for predicted in category_id_df.category_id:
        for actual in category_id_df.category_id:
          if predicted != actual and conf_mat[actual, predicted] >= 20:
            print("'{}' predicted as '{}' : {} examples.".
      ↪format(id_to_category[actual],

      ↪id_to_category[predicted],

                                                          conf_mat[actual,↵
      ↪predicted]))
```

```
    display(df2.loc[indices_test[(y_test == actual) & (y_pred ==␣
↪predicted)]][['Product',
                                                                    ␣
↪'Consumer_complaint']])
    print('')
```

'Consumer Loan' predicted as 'Credit reporting, repair, or other' : 36 examples.

	Product	Consumer_complaint
2924625	Consumer Loan	In XX/XX/XXXX, I contacted the 3 credit bureau…
2333569	Consumer Loan	Made auto payment that cleared through my bank…
3851916	Consumer Loan	You guys responded the account is cclosed but …
850570	Consumer Loan	I am reporting The XXXXXXXX Family on behalf o…
856494	Consumer Loan	I never received this loan the company showing…
1922793	Consumer Loan	I was browsing the various options for product…
1857214	Consumer Loan	Two payments were made, the first on XX/XX/XXX…
535820	Consumer Loan	To Whom It May Concern : RE : ACCOUNT- XXXX Go…
1319584	Consumer Loan	On Friday XX/XX/2018 I spoke with Collections …
2449466	Consumer Loan	Ford Motor Credit is reporting a repossession …
248211	Consumer Loan	Harassing mail and told wouldnt put this on my…
3522805	Consumer Loan	Im unable to connect my account to a payment a…
1249076	Consumer Loan	I am inquiring about account numbers, which be…
434760	Consumer Loan	Navient /XXXX XXXX keeps reporting and damagin…
3658076	Consumer Loan	Navient is showing on my credit report several…
434789	Consumer Loan	I have not received confirmation not certified…
845664	Consumer Loan	I have reported identity theft and fraud to we…
301269	Consumer Loan	BMW financial is reporting my account late whe…
1839614	Consumer Loan	Hi my name XXXX XXXX XXXX XXXX i was XXXX on …
1435705	Consumer Loan	My student loans under one account with Nelnet…
1011896	Consumer Loan	I went onto Springleaf 's website to see if I …
2297916	Consumer Loan	I pulled my annual credit report and found a c…
528176	Consumer Loan	Well its been actually a few years this proble…
2323901	Consumer Loan	My complaint is against a finance company call…
3305448	Consumer Loan	I submitted my complaint XX/XX/XXXX about the …
706482	Consumer Loan	In XXXX, XXXX, the State of XXXX Department Of…
2792301	Consumer Loan	I have reached out to have payments deferred. …
1926029	Consumer Loan	I attended XXXX XXXX as well as XXXX from XXX…
89995	Consumer Loan	I have a student loan through AES, they report…
349280	Consumer Loan	I have previously reported the following XXXX …
2270704	Consumer Loan	I did not authorize any car dealership to do a…
3515324	Consumer Loan	I sent a dispute letter to Toyota Financial re…
1023922	Consumer Loan	On Monday XX/XX/2021 at XXXX (EST), I receiv…
1612323	Consumer Loan	This was not sent is error i am a consumer by …
68378	Consumer Loan	I purchased my auto in XXXX. Made every paymen…
1473782	Consumer Loan	I submitted a dispute with Exeter finance XX/X…

'Credit card or prepaid card' predicted as 'Credit reporting, repair, or other'

: 25 examples.

```
                                       Product  \
1054470  Credit card or prepaid card
2218159  Credit card or prepaid card
893272   Credit card or prepaid card
1289219  Credit card or prepaid card
2659509  Credit card or prepaid card
392833   Credit card or prepaid card
1388833  Credit card or prepaid card
3752579  Credit card or prepaid card
3596219  Credit card or prepaid card
2787125  Credit card or prepaid card
1924550  Credit card or prepaid card
3751542  Credit card or prepaid card
591676   Credit card or prepaid card
2209192  Credit card or prepaid card
2912346  Credit card or prepaid card
3855866  Credit card or prepaid card
209573   Credit card or prepaid card
862947   Credit card or prepaid card
721487   Credit card or prepaid card
707907   Credit card or prepaid card
2424774  Credit card or prepaid card
935655   Credit card or prepaid card
1006622  Credit card or prepaid card
1599838  Credit card or prepaid card
884893   Credit card or prepaid card


                                              Consumer_complaint
1054470  So many issues with this account I have been i…
2218159  My previous claim XXXX, against creditor USAA …
893272   For months I have not been able to access this…
1289219  My separated husband has opened credit cards i…
2659509  See the attached documents. I want the bureau …
392833   JP Morgan Chase has decided to decrease my cre…
1388833  I have been an exclusive Apple device owner & …
3752579  I went to the Conns store to purchase applianc…
3596219  I received a letter ( XX/XX/19 ) from WFFinanc…
2787125  There a Home Depot card in my name and made me…
1924550  On XX/XX/XXXX I applied for an offer that incl…
3751542  On XX/XX/2023 SHELL ( Citibank / Citgroup ) cl…
591676   Inaccurate information being reported on my re…
2209192  I have never missed a payment and I am very di…
2912346  Discover Bank is inaccurately reporting late '…
3855866  I got an email fromMoneyLion stating I opened …
209573   I recently had an extremely frustrating experi…
862947   My nephew and I dined at XXXX  As with XXXX  t…
```

```
721487    I saw an XXXX ad for CareCredit suggesting tha…
707907    Credit karma just settled a class action lawsu…
2424774   I had XXXX recently and could not pay my payme…
935655    I have a XXXX XXXX XXXX store credit card that…
1006622   We set up auto pay for Bank of America Alaska …
1599838   discover bank credit card was declared in bank…
884893    Capital one has charge back away more charges …


'Debt collection' predicted as 'Credit reporting, repair, or other' : 83
examples.

                   Product                         Consumer_complaint
1868286  Debt collection  I sent XXXX XXXX XXXX a letter asking to verif…
2123455  Debt collection  To Whom It May Concern : I have recently been …
1262366  Debt collection  I am a victim of identity theft. I recently le…
2678712  Debt collection  Dear Support Representative, I just discovered…
3882843  Debt collection  NEVER HAD A LEGAL OBLIGATION OR APPLY FOR A LI…
…              …                                              …
1775276  Debt collection  Enhanced Recovery placed an account from XXXX …
2874855  Debt collection  My name is  XXXX   XXXX , SSN   XXXX  DOB  XX…
1808825  Debt collection  Dear Consumer Financial Credit Protection Bure…
1499117  Debt collection  The credit bureaus refused to provide you thei…
2710554  Debt collection  I have Nissan Financial on my credit for a veh…


[83 rows x 2 columns]


'Credit reporting, repair, or other' predicted as 'Consumer Loan' : 21 examples.

                                  Product  \
1586860  Credit reporting, repair, or other
3776992  Credit reporting, repair, or other
2617093  Credit reporting, repair, or other
282382   Credit reporting, repair, or other
1489921  Credit reporting, repair, or other
3248696  Credit reporting, repair, or other
1272135  Credit reporting, repair, or other
2069144  Credit reporting, repair, or other
160985   Credit reporting, repair, or other
250793   Credit reporting, repair, or other
735760   Credit reporting, repair, or other
1427474  Credit reporting, repair, or other
3155816  Credit reporting, repair, or other
2195654  Credit reporting, repair, or other
822275   Credit reporting, repair, or other
1405767  Credit reporting, repair, or other
232301   Credit reporting, repair, or other
1441492  Credit reporting, repair, or other
1116468  Credit reporting, repair, or other
```

```
204155   Credit reporting, repair, or other
621523   Credit reporting, repair, or other


                                         Consumer_complaint
1586860  I went to cash my payroll check at XXXX and it…
3776992  I opened an account with West Creek Financial …
2617093  I received a call from a student loan company …
282382   I have contacted Navient about the account rep…
1489921  I applied for student loan forgiveness for 4 s…
3248696  Wrong addresses. precious wong name listings. …
1272135  I did a loan consolidation in XX/XX/XXXX with …
2069144  Westlake Financial Services has placed false &…
160985   Sometime around the end of XXXX, beginning of …
250793   On XX/XX/2021, never applied for any car loans…
735760   I have Navient student loans that are private …
1427474  Navient has been reporting on my credit ( co-s…
3155816  Exeter finance lender money for car sum of app…
2195654  I opened a US bank account in late XXXX that m…
822275   At the beginning of XX/XX/2018, I had finished…
1405767  On XX/XX/2021, my credit was ran without my pe…
232301   Non-Compliance of FRCA/GLBA On the evening of …
1441492  I just signed up with Clear One Advantage, a c…
1116468  My credit score from the three major reporting…
204155   I don't have an XXXX loan with XXXX XXXX  I am…
621523   They added another name to my Trans union acco…


'Credit card or prepaid card' predicted as 'Consumer Loan' : 23 examples.

                            Product  \
530410   Credit card or prepaid card
1478167  Credit card or prepaid card
1020831  Credit card or prepaid card
3471150  Credit card or prepaid card
4017371  Credit card or prepaid card
3604548  Credit card or prepaid card
1644913  Credit card or prepaid card
1549715  Credit card or prepaid card
3933004  Credit card or prepaid card
3022306  Credit card or prepaid card
484555   Credit card or prepaid card
616357   Credit card or prepaid card
3839758  Credit card or prepaid card
1225155  Credit card or prepaid card
144876   Credit card or prepaid card
1377002  Credit card or prepaid card
176908   Credit card or prepaid card
3514429  Credit card or prepaid card
3391522  Credit card or prepaid card
```

```
3949860  Credit card or prepaid card
1680752  Credit card or prepaid card
2306795  Credit card or prepaid card
1139198  Credit card or prepaid card


                                        Consumer_complaint
530410   - Approved for BBVA ClearPoints card on XX/XX/…
1478167  I went around XXXX XXXX time on XX/XX/2022 to …
1020831  On XX/XX/XXXX wife passed, on or about XX/XX/X…
3471150  I went to XXXX and did a rental with budget ca…
4017371  Executive Summary. This is a breach of contrac…
3604548  I paid for a watch with 2 day shipping so I co…
1644913  I I have a checking account US Bank and my acc…
1549715  I have an open credit account with Synchrony B…
3933004  My account was closed after informing the bank…
3022306  In XXXX of XXXX I was issued a Bank of America…
484555   I have had a prepaid account with this company…
616357   On XX/XX/2020 I opened a Chase Total Checking …
3839758  I disputed transactions on my account and was …
1225155  My cashapp was charged a large amount that I d…
144876   I have filed a previous complaint but jhave NO…
1377002  My identity was stolen, and Bluebird American …
176908   XXXX switched from XXXX to XXXX and chose Citi…
3514429  Used postal mail to correspond about an invest…
3391522  I purchased tickets to a XXXX XXXX XXXX game f…
3949860  Today is XX/XX/XXXX and I call capital one thr…
1680752  When my husband and I needed to replace our ho…
2306795  I have a {$2000.00} + on child support I have …
1139198  Chase has made it difficult for consumers by t…


'Debt collection' predicted as 'Consumer Loan' : 21 examples.
              Product                        Consumer_complaint
1478128  Debt collection  i was renting spot from a man to open a XXXX i…
3230112  Debt collection  MY IDENTITY WAS STOLEN FROM ME. PNC CARD WAS N…
594811   Debt collection  XX/XX/XXXX. email to company XXXX XXXX XXXX XX…
3501817  Debt collection  XX/XX/2020 WE my wife and I bought a set of so…
674212   Debt collection  I received a notice on XX/XX/XXXX stating that…
2608566  Debt collection  XXXX XX/XX/2020, both consumers entered into a…
3402892  Debt collection  XXXX XXXX XXXX XXXX XXXX \nSearch XXXX  Multi-…
1417966  Debt collection  VRS debited checking account without approval …
3460196  Debt collection  Back in XXXX, I reached out to Klarna about an…
2463837  Debt collection  Back in XXXX XXXX XXXX I had repeatedly contac…
3227368  Debt collection  I believe Patient Accounts Bureau is attemptin…
1248144  Debt collection  They claim I owe {$290.00} and I have received…
1275523  Debt collection  rented a truck from budget, had a blowout, bla…
1693432  Debt collection  I opened a XXXX XXXX  account in XX/XX/XXXX an…
2042793  Debt collection  on XX/XX/XXXX, both consumers entered into a p…
```

```
1083665  Debt collection  On XX/XX/2021 XXXX XXXX XXXX alleges that I ma…
670042   Debt collection  On XX/XX/2017 I was informed by XXXX XXXX XXXX…
1617790  Debt collection  I opened an account with XXXX ( Commenity Bank…
300422   Debt collection  I received a email from Navient to call regard…
3790692  Debt collection  I work for the government and was part of the …
2093681  Debt collection  My personal student loan from XXXX College was…
```

'Credit reporting, repair, or other' predicted as 'Debt collection' : 44
examples.

```
                                 Product  \
748589   Credit reporting, repair, or other
2985487  Credit reporting, repair, or other
549046   Credit reporting, repair, or other
1944842  Credit reporting, repair, or other
3631206  Credit reporting, repair, or other
2839015  Credit reporting, repair, or other
3467287  Credit reporting, repair, or other
3599423  Credit reporting, repair, or other
863801   Credit reporting, repair, or other
3185153  Credit reporting, repair, or other
283642   Credit reporting, repair, or other
1161955  Credit reporting, repair, or other
1726839  Credit reporting, repair, or other
598490   Credit reporting, repair, or other
850587   Credit reporting, repair, or other
492070   Credit reporting, repair, or other
2637801  Credit reporting, repair, or other
1422796  Credit reporting, repair, or other
1762979  Credit reporting, repair, or other
330870   Credit reporting, repair, or other
712206   Credit reporting, repair, or other
1199239  Credit reporting, repair, or other
1451978  Credit reporting, repair, or other
395460   Credit reporting, repair, or other
3338129  Credit reporting, repair, or other
722504   Credit reporting, repair, or other
2802188  Credit reporting, repair, or other
514123   Credit reporting, repair, or other
285769   Credit reporting, repair, or other
2380083  Credit reporting, repair, or other
3699893  Credit reporting, repair, or other
3581361  Credit reporting, repair, or other
2662963  Credit reporting, repair, or other
3266042  Credit reporting, repair, or other
146681   Credit reporting, repair, or other
2071914  Credit reporting, repair, or other
3310827  Credit reporting, repair, or other
```

```
734251   Credit reporting, repair, or other
223375   Credit reporting, repair, or other
1836919  Credit reporting, repair, or other
1332139  Credit reporting, repair, or other
112754   Credit reporting, repair, or other
927890   Credit reporting, repair, or other
3137695  Credit reporting, repair, or other

                                    Consumer_complaint
748589   XXXX XXXX determined a loan that they had for …
2985487  I sent of letter of dispute asking about proof…
549046   I C Systems has negatively reported {$2600.00}…
1944842  I have disputed XXXX XXXX because this was acc…
3631206  Debt owed to the collection agency has been pa…
2839015  The company Portfolio Recovery sent a letter t…
3467287  On my XXXX credit report dated XX/XX/XXXX I ha…
3599423  On XXXX XXXX, 2017, Monterey Collections Sv. w…
863801   I received a release of IRS and Minnesota Reve…
3185153  Barclays Delaware creditor have some fraudulen…
283642   Consumer Finance, On a recent review of my cre…
1161955  On XX/XX/2022, LVNV Funding LLC added a deroga…
1726839  Ive contacted XXXX, Professional Placement Ser…
598490   XXXX XXXX XXXX was given a opportunity to pres…
850587   This is the 3 time I have disputed and filed a…
492070   Between XX/XX/2018 and XXXX XXXX XXXX XXXX XXX…
2637801  I have previously sent you a request to valida…
1422796  Contacted by Convergent Outsourcing , Inc., XX…
1762979  I do not have a signed contract nor any contra…
330870   I tried to run my credit to buy a house found …
712206   This complaint is regarding Global Payment Che…
1199239  I am asking you to verify this debt and where …
1451978  I disputes this account on XX/XX/2023. I have …
395460   XXXX and Transunion are reporting incorrectly …
3338129  PLEASE SEE THE ATTACHED LETTERS THAT WAS SENT …
722504   According to law if the bureau has incorrect i…
2802188  I was contacted by XXXX XXXX. and they submitt…
514123   Navy federal has sent my information to XXXX X…
285769   On XX/XX/2018 I mailed XXXXXXXX XXXX Card Paym…
2380083  Citibank has recently put a collection account…
3699893  I have multiple things on my credit report tha…
3581361  I have disputed this item with the credit repo…
2662963  I received a denial for credit notice from XXX…
3266042  I lost my wallet a year ago and now someone op…
146681   Is a amount of this Law Firm and I never sign …
2071914  I have a loan showing open XX/XX/XXXX with the…
3310827  XXXX XXXX XXXX XXXX inserted this entry on my …
734251   I have contacted LVNV Funding, XXXX, XXXX, and…
223375   Originally I contacted XXXX XXXX on XX/XX/2021…
```

```
1836919   Working with XXXX XXXX to remove. XXXX 's put …
1332139   I filed The Erronious nater of the Collection …
112754    I am continually being called on the telephone…
927890    I have repeatedly requested for documentation …
3137695   On XXXX XXXX a derogatory mark was added on my…
```

[26]:
```python
model.fit(features, labels)

N = 4
for Product, category_id in sorted(category_to_id.items()):
  indices = np.argsort(model.coef_[category_id])
  feature_names = np.array(tfidf.get_feature_names_out())[indices]
  unigrams = [v for v in reversed(feature_names) if len(v.split(' ')) == 1][:N]
  bigrams = [v for v in reversed(feature_names) if len(v.split(' ')) == 2][:N]
  print("\n==> '{}':".format(Product))
  print("  * Top unigrams: %s" %(', '.join(unigrams)))
  print("  * Top bigrams: %s" %(', '.join(bigrams)))
```

```
C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm\_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(


==> 'Consumer Loan':
  * Top unigrams: navient, money, loan, chime
  * Top bigrams: debit card, line credit, account charged, savings account

==> 'Credit card or prepaid card':
  * Top unigrams: card, capital, citi, synchrony
  * Top bigrams: balance transfer, paypal credit, best buy, american express

==> 'Credit reporting, repair, or other':
  * Top unigrams: equifax, experian, transunion, reporting
  * Top bigrams: xxxx xxxx, told company, delete account, payments account

==> 'Debt collection':
  * Top unigrams: debt, collection, calls, apartment
  * Top bigrams: loan took, portfolio recovery, xxxx say, investigation remove

==> 'Mortgage':
  * Top unigrams: mortgage, escrow, modification, home
  * Top bigrams: title company, credit like, escrow account, selene finance
```

7 MODEL SELECTION AND IMPLEMENTATION

```
[27]: X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                  test_size=0.25,
                                                  random_state = 0)

      tfidf = TfidfVectorizer(sublinear_tf=True, min_df=5,
                              ngram_range=(1, 2),
                              stop_words='english')

      fitted_vectorizer = tfidf.fit(X_train)
      tfidf_vectorizer_vectors = fitted_vectorizer.transform(X_train)

      model = LinearSVC().fit(tfidf_vectorizer_vectors, y_train)
```

C:\Users\sjaru\miniconda3\Lib\site-packages\sklearn\svm_classes.py:32:
FutureWarning: The default value of `dual` will change from `True` to `'auto'`
in 1.5. Set the value of `dual` explicitly to suppress the warning.
  warnings.warn(

7.0.1 PREDICTIONS

```
[28]: new_complaint_credit = """I noticed an error on my credit report. There is an
      ↪account that does not belong to me, and it's affecting my credit score."""
      print(model.predict(fitted_vectorizer.transform([new_complaint_credit])))
```

['Credit reporting, repair, or other']

```
[29]: complaint_credit_2 = """I found inaccuracies in my credit report, and I have
      ↪already submitted disputes to correct them. However, these errors are still
      ↪affecting my credit score."""
      print(model.predict(fitted_vectorizer.transform([complaint_credit_2])))
```

['Credit reporting, repair, or other']

```
[30]: new_complaint_debt = """I keep receiving calls from a debt collection agency
      ↪for a debt that I have already paid off. This is harassment, and I want it
      ↪to stop."""
      print(model.predict(fitted_vectorizer.transform([new_complaint_debt])))
```

['Debt collection']

```
[31]: complaint_debt_2 = """I've been receiving relentless calls from a debt
      ↪collector, even after I've explained that I cannot make payments at the
      ↪moment due to financial hardship."""
      print(model.predict(fitted_vectorizer.transform([complaint_debt_2])))
```

['Debt collection']

```
[32]: new_complaint_loan = """I applied for a consumer loan, but the interest rate
      ↪they offered is much higher than what was initially advertised. This is
      ↪misleading and unfair."""
      print(model.predict(fitted_vectorizer.transform([new_complaint_loan])))
```

['Consumer Loan']

```
[33]: new_complaint_loan_2 = """I recently applied for a consumer loan to purchase a
      ↪new car. The loan was approved,
                          but the interest rate they offered is much higher
      ↪than what I was initially quoted.
                          This is causing a significant financial burden, and
      ↪I believe this is unfair and deceptive lending."""
      print(model.predict(fitted_vectorizer.transform([new_complaint_loan_2])))
```

['Consumer Loan']

```
[34]: new_complaint_mortgage = """I've been trying to get a mortgage for my new home,
      ↪but the mortgage company has been delaying the process and requesting
      ↪unnecessary paperwork."""
      print(model.predict(fitted_vectorizer.transform([new_complaint_mortgage])))
```

['Mortgage']

```
[35]: complaint_mortgage_2 = """My mortgage lender has been repeatedly making errors
      ↪in processing my payments, which is affecting my credit score and causing
      ↪financial stress."""
      print(model.predict(fitted_vectorizer.transform([complaint_mortgage_2])))
```

['Credit reporting, repair, or other']