Multi – Class Classification On Consumer Complaint Dataset

(Consumer Complaints about various financial products and services)

Verzeo Major Project - ML07B17

Group Members:
- Devanshu Garg (gargdevanshu0804cse@gmail.com)
- Gaurav Verma (gverma0549@gmail.com)
- Ekagra Shukla (ekagrashukla2@gmail.com)
- Harsh Mohan Agarwal (harshmohanagarwal28@gmail.com)
- Abheet Singh (abheetsingh8@gmail.com)
- Ayushi Singh (ayushijadoun25@gmail.com)
- Harsh Porwal (harshporwal.hp@gmail.com)

Data Source: **https://catalog.data.gov/dataset/consumer-complaint-database**

Dataset which we used can be downloaded from the following link:
https://bit.ly/3iRsbOT

(Size of this dataset is very large, we have reduced it by cleaning and sampling it after performing some computations. The reduced dataset can be downloaded from the link : **https://bit.ly/3iQ7C5h**)

Link to GitHub Repository of this project:
https://github.com/gargdevanshu/ML07B17

Objective

Classify Consumer Complaints into Predefined Product Categories.

Classification Algorithms used:
- Linear Support Vector Machine (Linear SVC),
- Random Forest Classifier,
- Multinomial Naive Bayes,
- Logistic Regression,
- K Nearest Neighbors Classifier (KNN), and
- Decision Tree Classifier.

Summary

Q1. What are the most common unigrams and bigrams to each Product Category ?

Following are the top 5 most common words from each Product category :
(Refer .ipynb file for both **Code** and **Explanation**, used to get the following answer)

- Bank account or service
    - Unigrams : bank, scottrade, overdraft, branch, deposit
    - Bigrams : checking account, debit card, overdraft fee, scottrade bank, citigold checking
- Checking or savings account
    - Unigrams : overdraft, deposit, branch, bank, bonus
    - Bigrams : overdraft fee, checking account, saving account, debit card, direct deposit
- Consumer Loan
    - Unigrams : car, vehicle, ally, ford, dealership
    - Bigrams : auto loan, regional acceptance, car loan, ally financial, lien release
- Credit card or prepaid card
    - Unigrams : card, amex, express, citi, reward
    - Bigrams : credit card, america express, balance transfer, annual fee, late fee
- Credit reporting & repair services, other personal consumer reports
    - Unigrams : equifax, inquiry, report, experian, transition
    - Bigrams : credit report, credit bureau, hard inquiry, credit file, credit reporting
- Debt collection
    - Unigrams : debt, collection, collect, collector, recovery
    - Bigrams : collection agency, collect debt, debt collector, debt collection, portfolio recovery
- Money transfer, virtual currency, or money service
    - Unigrams : coinbase, pal, moneygram, wu, bitcoin
    - Bigrams : western union, coinbase account, money transfer, account coinbase, money gram
- Mortgage
    - Unigrams : mortgage, modification, escrow, ocwen, foreclosure
    - Bigrams : loan modification, mortgage payment, mortgage company, escrow account, short sale
- Other financial service
    - Unigrams : vast, productive, face, 5900, deliquency
    - Bigrams : check payable, 34 month, month service, verified time, inspection report
- Payday loan, title loan, or personal loan
    - Unigrams : payday, ace, loan, cashnet, borrowed
    - Bigrams : payday loan, ace cash, cash express, applied payday, loan illegal
- Student loan
    - Unigrams : navient, student, loan, school, repayment

- Bigrams : student loan, private loan, income based, loan forgiveness, loan navient
- Vehicle loan or lease
  - Unigrams : vehicle, car, honda, santander, lease
  - Bigrams : bow financial, honda financial, santander consumer, gap insurance, used vehicle

Q2. Which Machine Learning Algorithm suits best for this problem ?

The DataFrame showing the accuracy of various Machine Learning Algorithms implemented is present in Summary Section of .ipynb file.
(Refer .ipynb file for both **Code** and **Explanation**, used to get the following answer)

Best suited Algorithm for this problem is : **Linear SVC** (Based on Test Accuracy)
Best suited Algorithm for this problem is : **Decision Tree** (Based on Train Accuracy)

Q3. Which Product Category has maximum number of Consumer Complaints ?

Following is the list showing the various Product Categories along with Number of Consumer Complaints:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

 * Credit reporting & repair services, other personal consumer reports --> 4817
 * Debt collection --> 3389
 * Mortgage --> 1996
 * Credit card or prepaid card --> 1712
 * Student loan --> 830
 * Bank account or service --> 548
 * Checking or savings account --> 527
 * Consumer Loan --> 402
 * Money transfer, virtual currency, or money service --> 266
 * Payday loan, title loan, or personal loan --> 253
 * Vehicle loan or lease --> 250
 * Other financial service --> 10

Product Category having maximum number of Consumer Complaints is : **Credit reporting & repair services, other personal consumer reports**

Q4. Which Company has maximum number of Consumer Complaints ?

Following is the list showing the top 10 Companies along with Number of Consumer Complaints:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

Company
EQUIFAX, INC.   1483

Experian Information Solutions Inc.   1225
TRANSUNION INTERMEDIATE HOLDINGS, INC.   1196
WELLS FARGO & COMPANY   552
CITIBANK, N.A.   546
BANK OF AMERICA, NATIONAL ASSOCIATION   521
JPMORGAN CHASE & CO.   514
Navient Solutions, LLC.   486
CAPITAL ONE FINANCIAL CORPORATION   398
SYNCHRONY FINANCIAL   257
Name: Company, dtype: int64

Company having maximum number of Consumer Complaints is : **EQUIFAX, INC.**

Q5. What is the response of the companies to maximum number of Consumer Complaints ?

Following is the list showing the Company's Response to Complaints along with Number of Consumer Complaints with that Response:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

Closed with explanation   12081
Closed with non-monetary relief   1785
Closed with monetary relief   880
Closed   167
Untimely response   87
Name: Company_response_to_consumer, dtype: int64

Company's Response to maximum number of Consumer Complaints is : **Closed with explanation** without any monetory and non-monetary relief.

Q6. Which State has maximum number of Complainants ?

Following is the list showing the State along with Number of Complainants belonging to that State:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

State
CA   2025
FL   1493
TX   1454
GA   896
NY   836
IL   564
NJ   522
PA   515
OH   484
NC   463

Name: State, dtype: int64

State with Maximum Number of Complainants is : **CA (California)**

Q7. Which Year has maximum number of Consumer Complaints ?

Following is the list showing the Year along with Number of Consumer Complaints in that year:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

year
2018   4592
2017   4496
2016   3097
2015   2137
2019   678
Name: year, dtype: int64

Year having maximum number of Consumer Complaints is : **2018**

Q8. Which Year has maximum number of Consumer Complaints ?

Following is the list showing the Month along with Number of Consumer Complaints in that month:
(Refer .ipynb file for both **Code** and **Explanation** along with **Plot**, used to get the following answer)

month
09   1401
01   1390
10   1359
08   1343
07   1259
06   1259
05   1204
03   1185
12   1180
04   1166
02   1142
11   1112
Name: month, dtype: int64

Month having maximum number of Consumer Complaints is : **September (09)**