1. INTRODUCTION

Customer satisfaction plays an important role in any company's growth. Whenever a company renders services or products, it needs to always have a support team to tackle the complaints of the customer from time to time. This is very much important for any organization. There is an immediate need for the companies to have a good support team who can handle large customer complaints. As we know that, this is impossible to achieve only with the human force. We need to involve intelligent machines that can help in classifying the customer complaints and automatically connect the customers to the right customer care category. Enabling this with the help of NLP is possible.

In this project, we will be dealing with the same mechanism for the bank and finance-related customer care service to tackle customer complaints on daily basis.

1.1. EXISTING SYSTEM

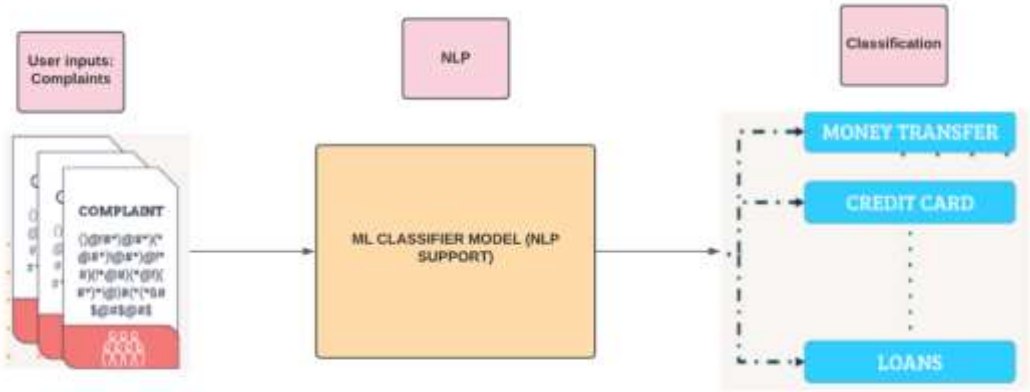
Customer complaint analysis utilizing various data mining approaches is a hot topic in academia right now. Recent studies have looked at and analyzed a wide range of service providers, such as airlines and banking services, to name a few. The value of this research is from the potential benefits that may be gained by understanding and identifying the variables that cause these complaints, which can then be used to propose appropriate solutions to improve these services.

Several Datamining processes are used over the customer complaints but in major cases, this process of only visualization does not automate the classification and connecting the right customer agent. No decision-making capacity for the existing systems in this regard.

In some research papers, they depicted customer complaints from mobile service providers are clustered using the K-means cluster analysis algorithm. The research focused on the creation of explanatory notes for complaint orders and the subsequent design of these orders. The study also performed statistical analysis on the complaints group, which is beneficial to the customer complaints group and clustering. But this clustering classification might not be feasible to apply when we know the particular classes of complaints that needed to be classified. As we need a supervised algorithm that would work efficiently for our application

1.2. PROPOSED SYSTEM

The proposed system mainly comprises the NLP concepts with the training of a machine learning algorithm that classifies the complaints into multiple classes. We



developed a supervised machine learning model that can take the input from the user and make a prediction of the category customer care in the bank to consult.

Fig 1.2.i. Proposed Mechanism Overview

We had used the following steps to complete this implementation of this project where we had performed Data analysis through visualization using EDA. Preprocessing of the data using NLTK and then training the ML model using this data with the help of multi-class classification algorithms.

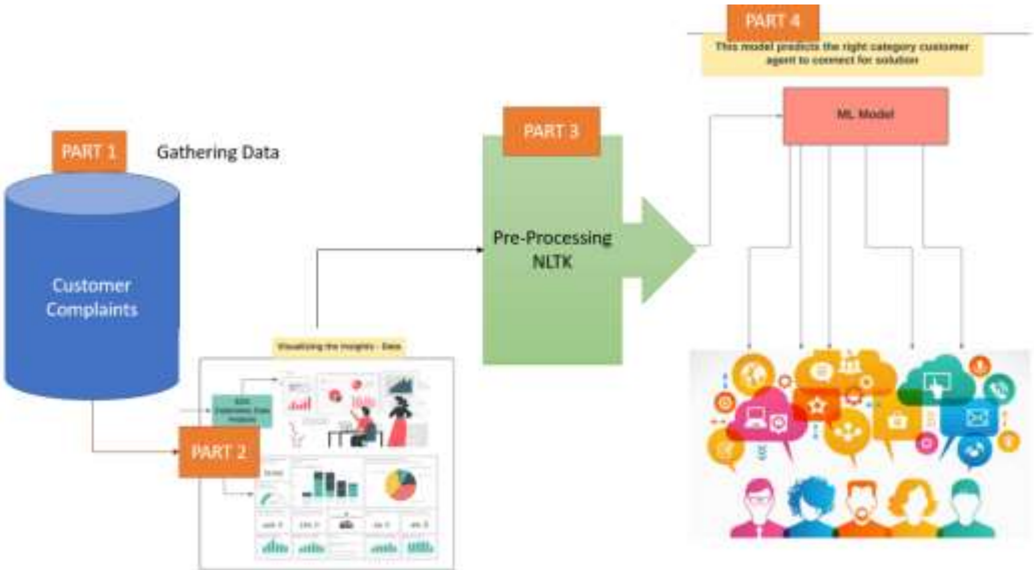


Fig 1.2.ii. Detailed Mechanism of the implementation of the proposed system

2. LITERATURE SURVEY

Processing customer complaints is a major challenge in the context of knowledge management technologies nowadays. Anon et al[proposed a model approach to improve automated processing of customer complaints. They have shown how communicative actions and attack links can be successfully modeled in terms of the graph-based representation provided by the notion of complaint scenario. They have also shown that their proposal for classifying com-plaint scenarios using supervised learning can be successfully applied, out per-forming the results obtained using the Complaint Engine platform when applying a keyword-based approach in which no attack links were taken into account for complaint classification. In this respect, the evaluation experiments using our dataset of formalized real-world complaints showed satisfactory performance.

In another study, Xu et al, used the K-means cluster analysis algorithm to cluster customer complaints from mobile service providers. The study referred to the formulation of explanatory notes for the complaint orders and then designing of these orders. The study conducted a statistical analysis of the complaints group as well, as this provides great support to the customer complaints group and their clustering.

According to Hsiao et al, an experimental study was conducted for a group of restaurants in Taiwan. The study aims to analyze customer complaints to process them, as well as to forecast and improve service quality. The decision tree methodology was integrated with Six Sigma analysis tools in this study and the results indicate a decrease in the level of customer complaints.

Yang presented a model for classifying customer complaints at a telecommunication company. Previous researches focused on decision support systems in handling complaints, while this study suggested using decision support systems based on (ER) evidential reasoning. The proposed model provides high performance in improving customer complaints handling.

Birim et al suggested a model for analyzing customer complaints. The research in focused on studying the impact of customer complaints on business performance in the airline sector. The study linked the fees and quality of service and important events such as economic contraction. These variables were analyzed to predict customer complaints that affect the purchase of tickets in the future.

Chugani et al focused on customer complaints analysis in several banks in more than one region. The study suggested a model for identifying and solving problems using data mining. Both cluster analysis and predictive modeling have been applied to find places where complaints are frequent as well as to find what caused them. This kind of analysis and study help to process problems, win customer loyalty and thus increase profits.

The purpose of this article is to examine and analyse data on consumer financial complaints in order to determine how many essentially identical complaints are associated with a similar bank, business, or product. Credit Uncovering, Mortgage, Debt Collection, Consumer Loan, and Banking Accounting grumbles cover both kinds of data. Using data mining technologies, cluster analysis, as well as farsighted showing, will be expanded in order to obtain substantial information regarding grumblings in certain areas of the country. Banks that accept client complaints will examine the data to see where the most grumbling is stored, what objects / organisations are generating the most problems, and other pertinent information. The methodology could aid banks in gaining a better understanding of the region and types of target mixups, as well as encouraging widespread buyer participation to enhance wages and profitability. Naïve Bayes mechanisms are proposed in this regard.

There are Knowledge base systems where the data is manually annotated. They have established bots with the help of KB’s. But this are static in nature as bots can only respond based on the predefined values without any intelligence.

3. DESIGN

3.1. REQUIREMENT SPECIFICATION:

SOFTWARE:

Coding Platform used: Google Colab

IDE: PyCharm 2020.3.3

Python Interpreter: Python 3.9

Libraries & versions:

numpy==1.18.5

pandas==0.25.3

python-dateutil==2.8.1

pytz==2020.1

scikit-learn==0.23.1

scipy==1.4.1

NLTK Libraries:

TfidfTranformer

3.2. DATA FLOW DIAGRAM

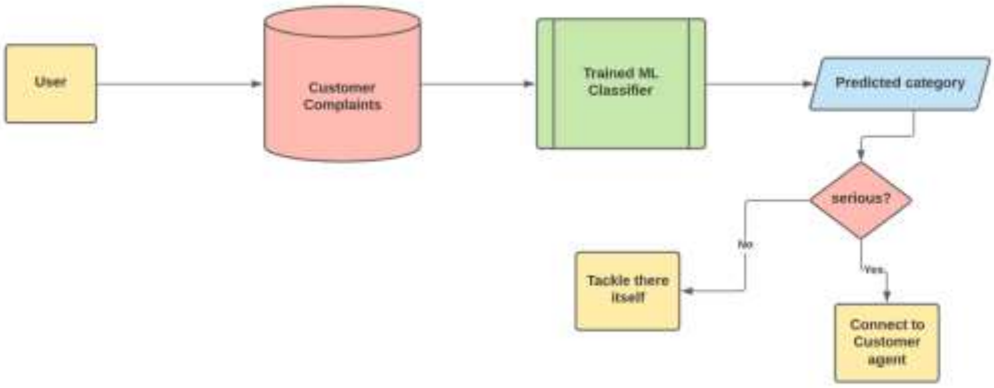


Fig 3.2.i. Data flow diagram of process

4. IMPLEMENTATION

Implementation of this proposed project can be done in below parts:

- Data Gathering
- Data Visualization
- Data Prep-processing using NLTK
- Training the ML classifier
- Testing the results
- Evaluating the model

In detailed the important steps include:

- Data Exploration (Understanding the dataset)
- Data (text) Pre-processing and the general steps one might take in an NLP project
- Model selection and why Micro-F1 is used as an evaluation metric
- Selected model final test and results

4.1. DATA GATHERING:

The data used to train this model may be accessed on the official site of the Consumer Complaint Database, along with an explanation of what each column signifies. It's essentially a tagged dataset of complaints received about financial products and services from the Consumer Financial Protection Bureau (CFPB) [9]. The dataset contains 1,437,716 rows of data, including a large number of rows containing null values.

Fig 4.1.i. Official site from where data is collected

Fig 4.1.ii. Consumer Complaint Database details

4.2. DATA REDUCTION & VISUALIZATION

- Dropping all rows that do not have Customer Complaint entries in them. – NULL values
- Subsetting Dataframe for Text Multi-Classification Problem
- Understanding/Cleaning Our Data
- Dropping the columns not needed.

```
[7] # Column Cleaning First. Columns are super nastily named
df.rename(columns={'Date received':'DATE_RECEIVED',
                   'Product':'PRODUCT',
                   'Sub-product':'SUB_PRODUCT',
                   'Issue':'ISSUE',
                   'Sub-issue':'SUB_ISSUE',
                   'Consumer complaint narrative':'CONSUMER_COMPLAINT_NARRATIVE',
                   'Company public response':'COMPANY_PUBLIC_RESPONSE',
                    'Company':'COMPANY',
                    'State':'STATE',
                    'ZIP code':'ZIP_CODE',
                    'Tags':'TAGS',
                    'Consumer consent provided?':'CONSUMER_CONSENT_PROVIDED',
                    'Subbmited via':'SUBMITTED_VIA',
                    'Date sent to company':'DATE_SENT_TO_COMPANY',
                    'Company response to consumer':'COMPANY_RESPONSE_TO_CONSUMER',
                    'Timely response?':'TIMELY_RESPONSE',
                    'Consumer disputed?':'CONSUMER_DISPUTED',
                    'Complaint ID':'COMPLAINT_ID'
                   }, inplace=True)
```

Fig 4.2.i. Cleaning the dataset

```
[11] df_product_and_complaint.head()
```

	PRODUCT	CONSUMER_COMPLAINT
1	Vehicle loan or lease	I contacted Ally on Friday XX/XX/XXXX after fa...
12	Credit reporting, credit repair services, or o...	Hello This complaint is against the three cred...
13	Credit reporting, credit repair services, or o...	I am a victim of Identity Theft & currently ha...
15	Credit reporting, credit repair services, or o...	Two accounts are still on my credit history af...
18	Credit reporting, credit repair services, or o...	Receiving daily telephone call (s) from XXXX...

Fig 4.2.ii. Dataset after dropping unnecessary columns

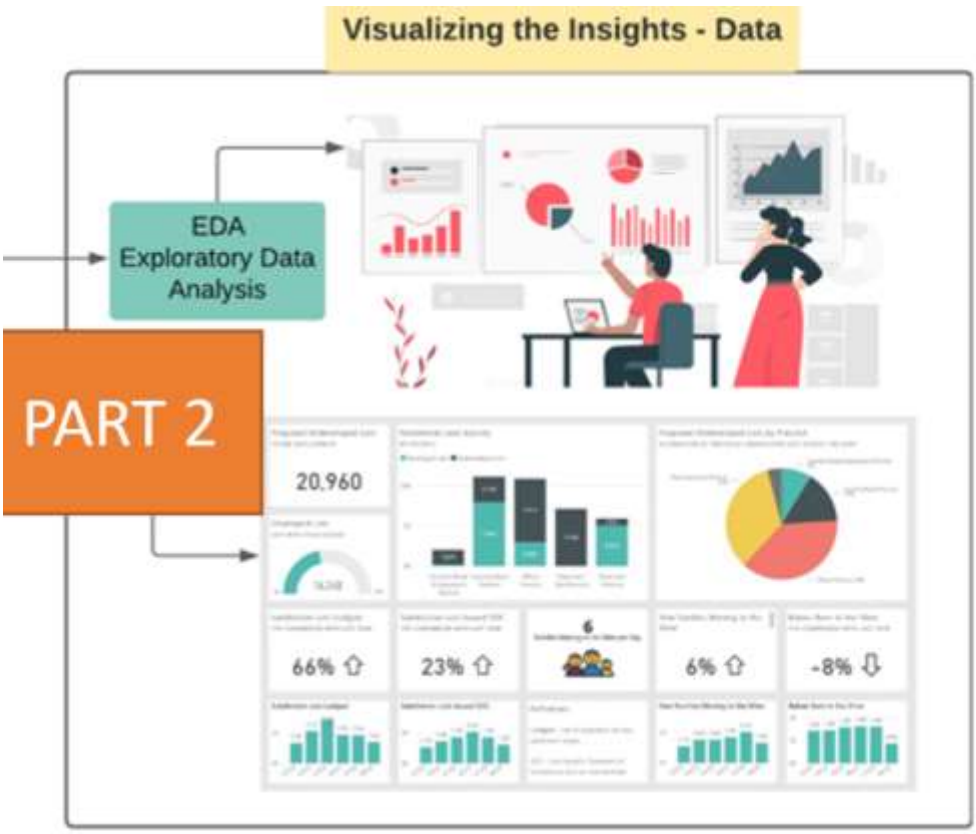


Fig 4.2.iii. We perform EDA on data

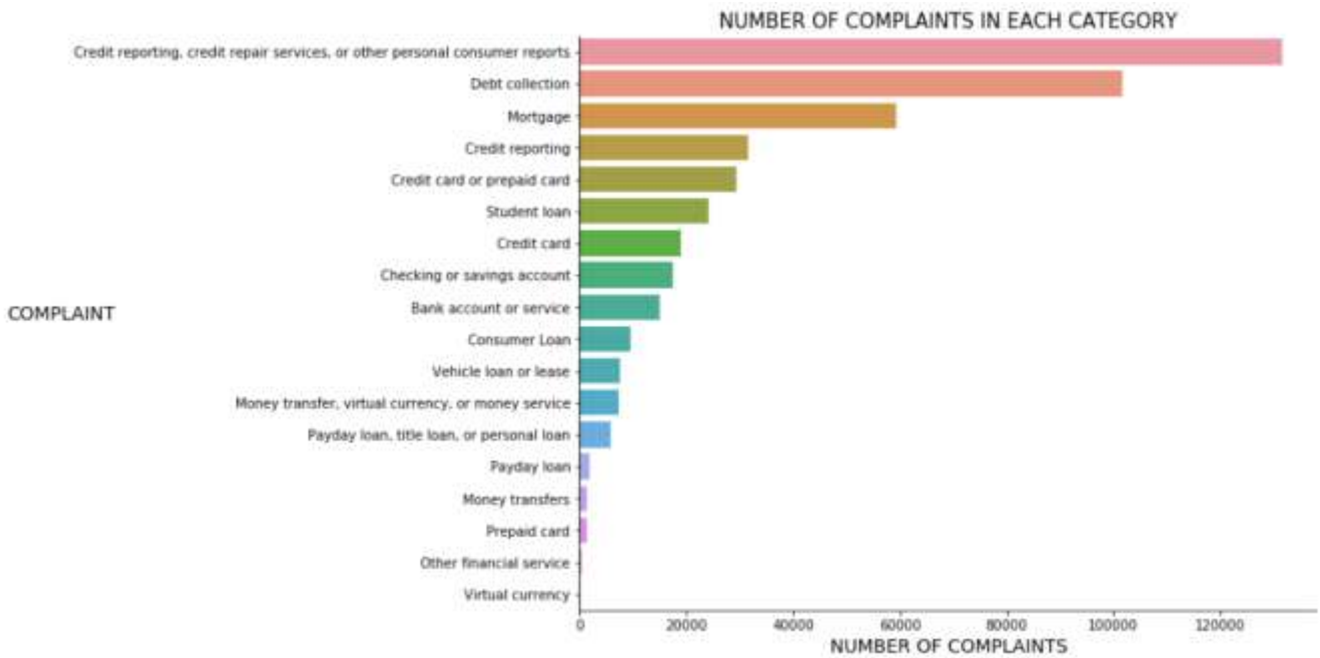


Fig 4.2.iv. After performing EDA on product column (complaint categories)

We observed that there are many categories that overlap with each other.

Like Credit card and credit card and prepaid card, Money transfers and money transfers, services. So we can remove these overlapping categories and generalize them so as to decrease the number of categories for classification.

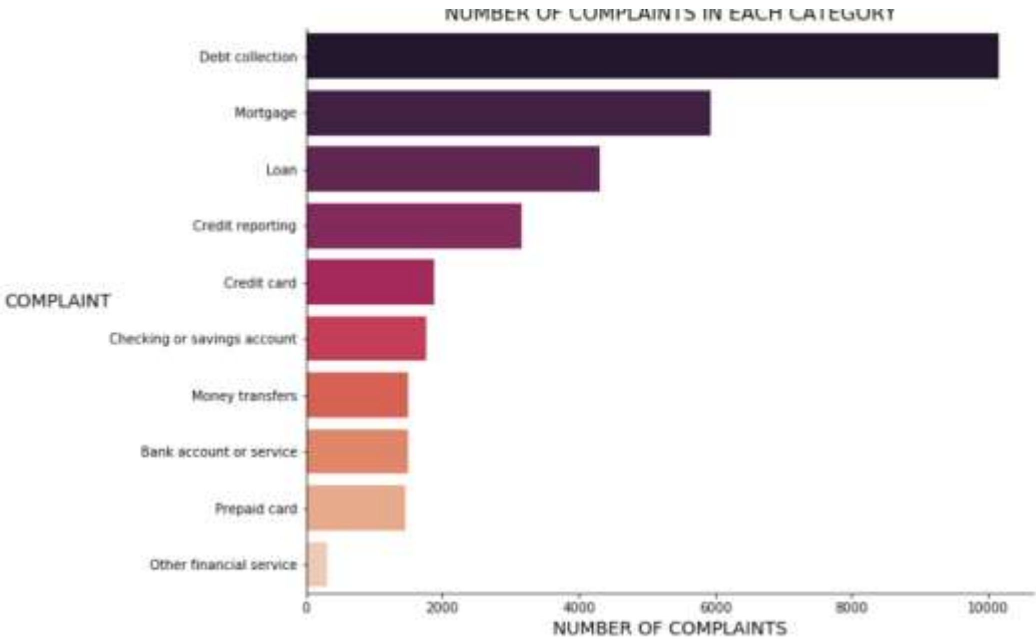


Fig 4.2.v. After reducing the overlapping categories

Reduced Dataset: 15196 rows, 3 columns, and 10 categories (result)

4.3. TEXT PREPROCESSING

We know that the machines cannot understand the high-level language (English) that we use. As our approach is computer-oriented, we should first convert them into machine-understandable.

Thus, to attain this we use NLTK (Natural Language Tool Kit).

Vectorization is a technique of converting the text into a machine-readable form where all the words are represented in form of vectors.

4.3.1. TF-IDF VECTORIZATION

Term Frequency: retrieves the frequency of the word in the current comment.

Inverse Document Frequency: retrieves how rare the word is across all comments.

Here, we want to highlight the words which are frequent in one particular comment but not across all comments. Thus, we use TfidfVectorizer here.

If a word is present in all comments, its IDF value will be low and vice versa.

Splitting of Dataset into Training data and test data:

```
X, y = df_product_and_complaint_reduced.CONSUMER_COMPLAINT, df_product_and_complaint_reduced.PRODUCT_ID
print('X shape:', X.shape, 'y shape:', y.shape)
```

```
X shape: (15196,) y shape: (15196,)
```

```
# Split the data into X and y data sets
X, y = df_product_and_complaint_reduced.CONSUMER_COMPLAINT, df_product_and_complaint_reduced.PRODUCT_ID
print('X shape:', X.shape, 'y shape:', y.shape)

# For text classification, ALWAYS split data first before vectorizing.
# This is because you don't want to cheat by having features (words) from the test data already being inside your train data
from sklearn.model_selection import train_test_split

X_train_val, X_test, y_train_val, y_test = train_test_split(X, y,
                                                            test_size=0.2,   # 80% train/cv, 20% test
                                                            stratify=y,
                                                            random_state=seed)
print('X_train', X_train_val.shape)
print('y_train', y_train_val.shape)
print('X_test', X_test.shape)
print('y_test', y_test.shape)
```

```
X shape: (15196,) y shape: (15196,)
X_train (12156,)
y_train (12156,)
X_test (3040,)
y_test (3040,)
```

Fig 4.3.1.i. Splitting of Dataset

```python
# Performing Text Pre-Processing

# Import tfidfVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer

# Text Preprocessing
# The text needs to be transformed to vectors so as the algorithms will be able make predictions.
# In this case it will be used the Term Frequency - Inverse Document Frequency (TFIDF) weight
# to evaluate how important A WORD is to A DOCUMENT in a COLLECTION OF DOCUMENTS.

# tfidf1 = 1-gram only.
tfidf1 = TfidfVectorizer(sublinear_tf=True, # set to true to scale the term frequency in logarithmic scale.
                         min_df=5,
                         stop_words='english')

X_train_val_tfidf1 = tfidf1.fit_transform(X_train_val).toarray()
X_test_tfidf1 = tfidf1.transform(X_test)

# tfidf2 = unigram and bigram
tfidf2 = TfidfVectorizer(sublinear_tf=True, # set to true to scale the term frequency in logarithmic scale.
                         min_df=5,
                         ngram_range=(1,2), # we consider unigrams and bigrams
                         stop_words='english')
X_train_val_tfidf2 = tfidf2.fit_transform(X_train_val).toarray()
X_test_tfidf2 = tfidf2.transform(X_test)
```

Fig 4.3.1.ii. TF-IDF Vectorization

N - Grams

Neighboring sequences of items (words, letters, or symbols) in a document.

E.g: Not satisfied with loan system.

Unigram: would be a list created of the following:

[‘Not’,’satisfied’,’with’,’loan’,’system’] --- (5 elements in a list)

Bigram (2-gram) [‘Not satisfied’,’satisfied with’,’with loan’,’loan system’] --- (4 elements in the list)

Stop words: frequent words in English that don’t provide any useful info.

4.3.1. Using GoogleNews Word2Vec300d

The word2vec tool accepts a text corpus as input and outputs word vectors. It learns vector representations of words after first constructing a vocabulary from the training text input [10]. Many natural language processing and machine learning systems can use the generated word vector file as a feature.

Word and phrase vectors that have been pre-trained

We're utilizing pre-trained vectors that have been trained on a subset of the Google News dataset (about 100 billion words)[11]. Three million words and phrases are represented by 300-

dimensional vectors in the model. As a result, we can state that we use GoogleNews Word2Vec200d for text preprocessing [12].

4.4. TRAINING MODEL

Multi-class Text Classification Algorithms are used:

- Multinomial Naive Bayes
- Logistic Regression
- Random Forest
- LinearSVC

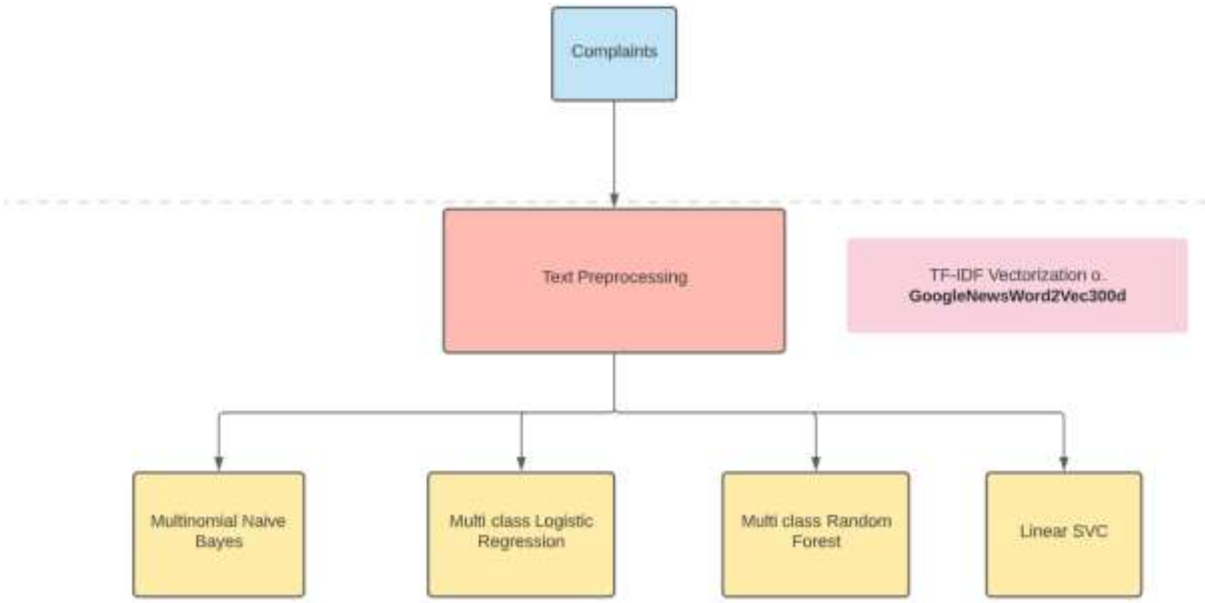


Fig 4.4.i. Applying different Algorithms

5. TESTING & RESULTS

5.1. MODEL EVALUATION TRAINED BY TF-IDF VECTORIZATION

results_cv_stratified_1gram

	Model	Accuracy	Macro Precision	Macro Recall	Macro F1score	Weighted Precision	Weighted Recall	Weighted F1	Time taken
0	MultinomialNB1	0.858094	0.352014	0.328769	0.334049	0.838411	0.858094	0.83823	28.483524

results_cv_stratified_2gram

	Model	Accuracy	Macro Precision	Macro Recall	Macro F1score	Weighted Precision	Weighted Recall	Weighted F1	Time taken
0	MultinomialNB2	0.825682	0.348484	0.302296	0.307413	0.817254	0.825682	0.794479	141.008026

Fig 5.1.i. Performance of Multinomial Naïve Bayes algorithm

results_cv_straitified_1gram

	index	Model	Accuracy	Macro Precision	Macro Recall	Macro F1score	Weighted Precision	Weighted Recall	Weighted F1	Time taken
0	0	MultinomialNB1	0.858094	0.352014	0.328769	0.334049	0.838411	0.858094	0.838230	30.445922
1	0	GaussianNB1	0.555693	0.252018	0.257251	0.244527	0.604821	0.555693	0.562513	108.363672
2	0	LogisticRegression1	0.904573	0.435528	0.378127	0.384980	0.883308	0.904573	0.891614	2522.540348
3	0	RandomForest1	0.566222	0.326971	0.150526	0.145625	0.711326	0.566222	0.471929	228.233541
4	0	LinearSVC1	0.908768	0.743714	0.493561	0.547539	0.902041	0.908768	0.901855	37.112729

Fig 5.1.ii. Performance of all Algorithms (using TF-IDF Vectorizer)

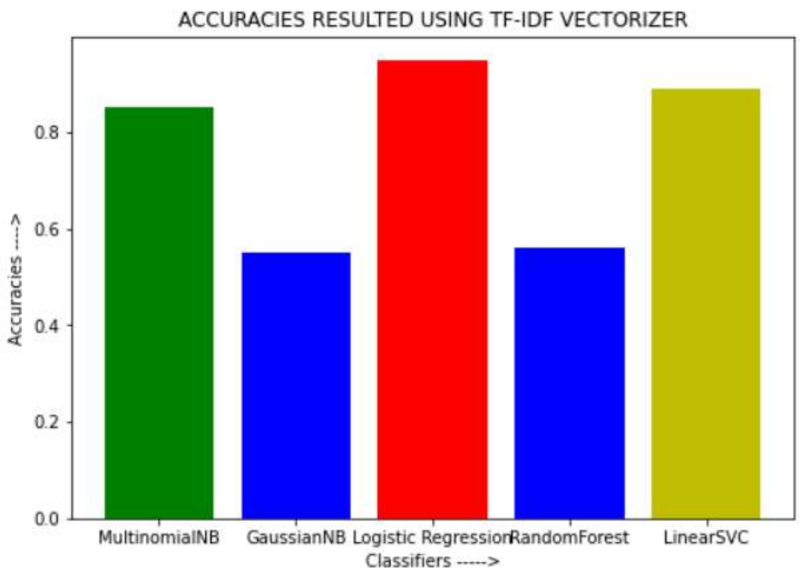


Fig 5.1.iii. Higher accuracy obtained by Multi-class Logistic regression

5.2. USING GOOGLENEWS WORD2VEC300D

	index	Model	Accuracy	Macro Precision	Macro Recall	Macro F1score	Weighted Precision	Weighted Recall	Weighted F1	Time taken
0	0	GaussianNB_word2vec300d	0.462022	0.440222	0.453180	0.420936	0.577487	0.462022	0.492835	11.035332
1	0	LogisticRegression_word2vec300d	0.744963	0.630964	0.585320	0.600587	0.729354	0.744963	0.731291	356.777384
2	0	RandomForest_word2vec300d	0.475678	0.438624	0.199697	0.174722	0.512870	0.475678	0.363007	474.810438
3	0	LinearSVC_word2vec300d	0.763983	0.697031	0.619604	0.624424	0.753017	0.763983	0.751898	474.778550

Fig 5.2.i. Performance of Algorithms using (Word2Vec)

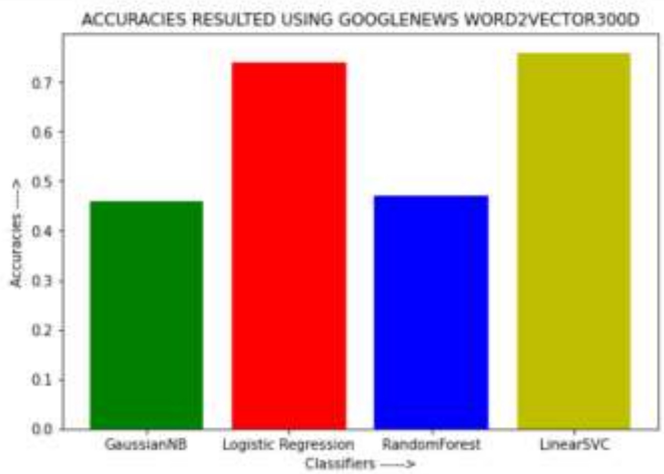


Fig 5.2.ii. Higher accuracy obtained by LinearSVC

5.3. PREDICTION

Hence, we got higher accuracy in case of Multi-class Logistic Regression using TF-IDF Vectorization = 90% of accuracy.

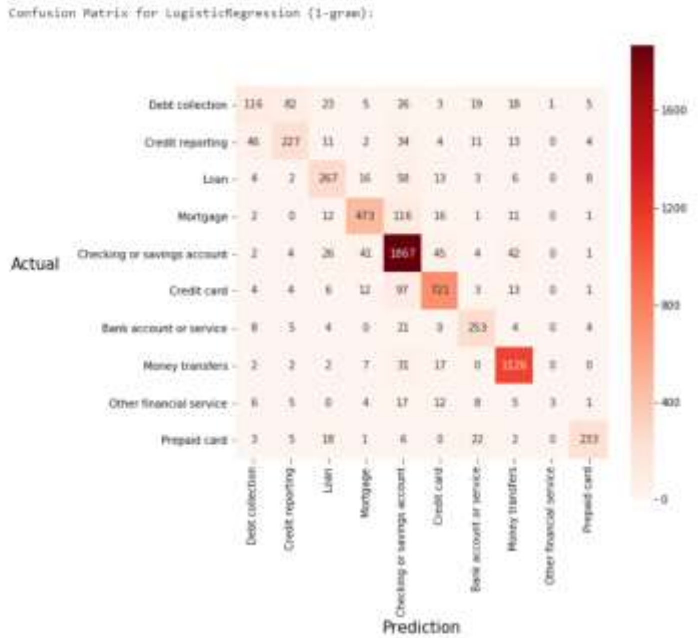


Fig 5.3.i. Confusion Matrix

```
Omg where is my money?????? AOSINIONSAD WHY DID YOU GUYS TAKE MY MONEY AWAY. SAOIDNSIOADNOIAODNNIOASDNSADNOSDNOASDN
I TRANSFERRED IT LAST NIGHT AND I WOKE UP TO NOTHING IN!!!!!! MY BANK ACCOUNT. HELP PLEASE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
I NEED THE MONEY OR ELSE I WILL BE HUNTED BY LOAN SHARKS!!!!!!!!
YOU *(#&$) PEOPLE HAVE NO RIGHT TO DO THIS TO ME!!!! I NEED MY MONEY!!!!

Money tranfers: 35.25%
Loan: 18.82%
Mortgage: 12.65%
Bank account or service: 9.88%
Checking or saving account: 8.94%
Prepaid card: 5.66%
Other financial service: 3.59%
```

Fig 5.3.ii. Prediction on a sample complaint

6. CONCLUSION

As we got high accuracy of 90% with the help of Logistic Regression of Multi-class for the bank related dataset where we can categorize complaints and connect to the specific customer agent respectively such that the customer problems can be tackled easily on time. This ensures customer satisfaction which indirectly helps in the growth of the business. We can develop an application that can take the recordings of the complaints in customer care. Then our proposed model can be applied which helps in classifying the complaints into various categories and connect the customer service who can reach out to the customer at right time to tackle their problems. Hence, tackling customer support made easy using the NLP model proposed.The scope of the project is to run the proposed system model in the backend of the financial customer service consultancies complaints web portals. Voice can be enabled for easy use.

7. REFERENCES

[1] Xu J, Zhao J, Zhao N, Xue C, Fan L, Qi Z, Wei Q. “The Research and Construction of Complaint Orders Classification Corpus in Mobile Customer Service”. In CCF International Conference on Natural Language Processing and Chinese Computing 2018 Aug 26 (pp. 351-361). Springer, Cham.

[2] Hsiao YH, Chen LF, Choy YL, Su CT. “A novel framework for customer complaint management”. The Service Industries Journal. 2016 Oct 25;36(13-14):675-98.

[3] Yang Y, Xu DL, Yang JB, Chen YW. “An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications”. Knowledge-Based Systems. 2018 Sep 25.

[4] Birim S, Anitsal MM, Anitsal İ. “A MODEL OF BUSINESS PERFORMANCE IN THE US AIRLINE INDUSTRY: HOW CUSTOMER COMPLAINTS PREDICT THE PERFORMANCE?”. Business Studies Journal. 2016 Jul 1;8(2).

[5] Chugani S, Govinda K, Ramasubbareddy S. “Data Analysis of Consumer Complaints in Banking Industry using Hybrid Clustering”. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) 2018 Feb 15 (pp. 74-78). IEEE.

[6] Anon,(n.d).[online]Available file:///C:/Users/HP/Downloads/MTA_CustomerFeedbackRIghtNowData_Overview%20(2).pdf [Accessed 20 Apr. 2017].

[7] Wang, Y.; Tseng, M.M. ―A Naïve Bayes approach to map customer requirements to product variants.‖ J. Intell. Manuf. 2015, 26, 501–509.

[8] Lara et al. (2018). Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police. Knowledge-Based Systems.

[9] https://cfpb.github.io/api/ccdb/fields.html

[10] https://code.google.com/archive/p/word2vec/

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.

[12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS, 2013.

[13] Gartner. (2019). Retrieved from Gartner Identifies Four Imperatives That Customer Service and Support Leaders Must Address to Deliver Digital Customer Service in 2019: https://www.gartner.com/en/newsroom/press-releases/2019-02-11-gartner-identifies-four-imperatives-t hat-customer-ser

[14] Brock, C. et al. (2013). Satisfaction with complaint handling: A replication study on its determinants in a business-to-business context. International Journal of Research in Marketing

[15] Coussement et al. (2009). Improving customer complaint management by automatic email classification using linguistic style features as predictors. Decision Support Systems, 870-882.