**Financial Complaint Text Classification**

Sameer Kulkarni   
Computer science and Engineering  
PES UniversityBengaluru, India

sameer6kulkarni@gmail.com

Sumukha M K   
Computer science and Engineering  
PES UniversityBengaluru, India

sumukhaganesh@gmail.com

Sarthak S   
Computer science and Engineering  
PES UniversityBengaluru, India

sarthakshivaraju@gmail.com

Sanket Kattimani   
Computer science and Engineering  
PES UniversityBengaluru, India

sanketskattimani150@gmail.com

Ashwini M Joshi   
Project Guide  
PES UniversityBengaluru, India

ashwinimjoshi@pes.edu

***Abstract***—***Text classification is one of the most interesting yet challenging tasks in Natural Language Processing. It has become more useful in the field of Financial complaint classification. Banks usually get many complaints through letters, social media or through the bank's website. Sometimes people address the wrong department while complaining and manual classification of complaints for banks also becomes a tedious task. So correct classification of complaints is necessary for better user experience and faster complaint solution.***

***Through this project we are trying to attain best results for our complaint classification task by comparing various Machine Learning (ML) models, Deep Learning (DL) models and Ensemble methods on basis of accuracy and time and applying the one which best suits the requirement.We are using data pre-processing methods like data augmentation, lemmatization etc and on top of that TF-IDF and Word2Vec methods for ML and DL models respectively.***

***Keywords—Finance, Complaint, Classification, Machine Learning, Deep Learning, Ensemble, Data augmentation, Natural Language Processing.***

# Introduction

Financial Companies get many complaints regarding their service and they have to handle those complaints quickly and effectively.Reading many complaints for the customer support people and tagging them to route to their respective departments becomes hectic and also takes time to handle the complaints. Also, now big financial companies have made their complaint handling system online but require the customers themselves to tag the department to which the complaint should be routed. There are high chances of customers choosing the wrong department while tagging themselves due to lack of knowledge about the departments in financial companies and there are also chances of them choosing a random department since they are only interested in writing their complaint and not filling other details. Hence, an Automated Financial Complaint Classification System becomes important that automatically tags incoming complaints by using Artificial Intelligence.

In this paper we are trying to attain best results for our complaint classification task by comparing various Machine Learning (ML) models, Deep Learning (DL) models and Ensemble methods on the basis of accuracy and time and applying the one which best suits the requirement. We also used data augmentation to deal with data imbalance and improve accuracy.

# Literature Review

We reviewed some research papers as part of a literature survey. In Complaint Classification using Word2Vec Model [1] they used GRU and MLP along with adam optimizer and Word2Vec for embedding layers. They got accuracy of 85%.Here they didn’t describe how they handled class imbalance problems. In the Complaint Text Classification Model Based on Character-level Convolutional Network [2] they used character level CNN along with NER. In Developing an LSTM based Classification Model of IndiHome Customer Feedbacks [3] they compared Naive bayes with LSTM using TF-IDF and found that LSTM was performing better than NB. In Automatic Complaint Classification System Using Classifier Ensembles [4] researchers compared ML models like SVM,NB etc with ensemble methods and got accuracy of 81% they represented data as BOW(bag of words) and applied models. TF-IDF could have resulted in better accuracy. In Active Learning SVM Classification Algorithm for Complaints Management Process Automatization [5] they showed a new method of SVM which needs less data yet gives better results than Decision Trees. In Bank CustomerComplaints Analysis Using Natural Language Processing and Data Mining [6] they used unsupervised learning method LDA(*Linear Discriminant Analysis)* for classification. InSentiment Classification of Indian Banks’ Customer Complaints [7] they used different ML models and found that Random forest and Naive bayes are the one which performed better that rest.

# methodology used

1. *Data*

For our analysis, we used a dataset from the Consumer Financial Protection Bureau government website. We took the latest complaints starting from last year for our analysis. All the complaints description is present in the “Consumer\_complaint\_narrative” column and the category of complaint is present in the “Product” column. There are 9 categories in total in the Product column.

Since the dataset was highly imbalanced , we handled class imbalance problems by undersampling majority classes. For increasing data for minority classes, we used Data Augmentation using the technique of Synonym Replacement by replacing random six words with its synonyms for each complaint. Using these techniques, 20000 rows from each class are taken for further analysis.

1. *Exploratory Data Analysis*

We visualized our text data by finding the most frequent words present in each category by Word Cloud Visualization.

1. *Text pre-processing*

Initially all words are converted to lower case. Then punctuations and digits are removed. All common words like the,and which don’t have significant meaning represented by stop words are removed. All words are then converted to base form by applying lemmatization. Finally confidential information represented by x's are removed to clean the complaints text.

1. *Machine Learning and Model Building*

We used Term Frequency – Inverse Document Frequency technique for word vector representations. Then, we split the data into train and test data as 70% and 30% of data. Five Machine Learning algorithms are implemented and compared.

The accuracy of the model obtained by training using unigrams and bi-grams are listed in the table below

| Sl.No. | Model | Unigram Accuracy | Bigram Accuracy |
| --- | --- | --- | --- |
| 1 | Multinomial Naïve Bayes | 79.99% | 82.94% |
| 2 | Decision Tree | 71.33% | 65.35% |
| 3 | Linear SVM | 83.51% | 85.35% |
| 4 | Logistic Regression | 83.03% | 81.71% |
| 5 | K-nearest Neighbors | 73.76% | 69.10% |

Bi-gram version of Linear SVM gave a good accuracy of 85.35% among all models, but increased the model complexity and time to process the complaint.

1. *Ensemble Learning Model Building and Testing*

Two ensemble learning models namely Random Forest which is a bagging based model and XG Boost which is a boosting based model were built and the results obtained are shown below

| Sl.No. | Model | Accuracy with 100 estimators | Accuracy with 500 estimators |
| --- | --- | --- | --- |
| 1 | Random Forest | 83.83% | 84.78% |
| 2 | XG Boost | 82.37% | 84.68% |

As in the case of Bi-gram version of Linear SVM, here also using a larger number of estimators gives better accuracy but increases model complexity.

1. *Deep Learning Model Building and Testing*
2. *Long Short Term Memory (LSTM):* LSTM is a variant of RNN Model. It can handle long term dependencies better than vanilla RNN model and hence very useful for various NLP tasks. Maximum of 400 words from each complaint is taken and an embedding dimension of 100 is chosen along with padding to make the length of the complaint equal. Train test ratio of 80:20 along with a validation split of 10 percent is chosen. The number of epochs set is 9 along with a batch size of 64. Embedding layer is used between input and LSTM layer. Spatial Dropout of 20 percent is used. Next, a LSTM layer of 128 units is used with dropout and recurrent dropout percentage of 20 to avoid overfitting. Final layer consists of a Dense Layer of 9 neurons since there are nine classes and softmax activation function is being used since it is a multiclass text classification problem. Loss function used is categorical cross entropy along with Adam optimizer. Early Stopping with a patience factor of 2 is used which means the model will stop when a maximum of two epochs show no improvement in the validation set. Min\_delta is set to 0.0001 which means the minimum improvement should be greater than 0.0001 for it to be considered an improvement in Early Stopping. The model stopped running after 7 epochs due to Early Stopping. Training accuracy of 92.11% and validation set accuracy of 84.02% is obtained after 7 epochs. Testing set accuracy obtained is 83.72 %.
3. *Bi-Directional LSTM:-* This model handles better dependencies than vanilla LSTM Model since it reads input from both left and right sides. Testing Accuracy obtained is 84.13%.
4. *Word level Convolutional Neural Network:-* Two Convolutional layers with 128 filters each have been used with kernel size of 5. Activation function used in convolutional layers in relu. Convolutional layers are followed by Max pooling layer with pool size of 5 along with dropout rate of 30%. The rest of the model is trained similar to the LSTM Model. Testing accuracy achieved is 84.25%.
5. *Voting Classifier:-*

We then implemented a hard voting classification model using 3 models which gave good accuracy. We used Random Forest, XG boost and Linear SVM with bi-grams and got accuracy of 85.91%.

Bi-gram Linear SVM, Random Forest and XG Boost Hard Voting Classification model gave better accuracy but models were complex and took much time to process a complaint in real life.

To solve this issue we sampled 1,00,000 complaints from dataset and built a Voting classifier model as earlier but this time we used Multinomial Naive Bayes, Linear SVM and Logistic Regression as they were giving good accuracy as individual models and processing time also very quick.

The Hard Voting Classification model gave accuracy of 82.46% which is a good accuracy and time taken is also very less.

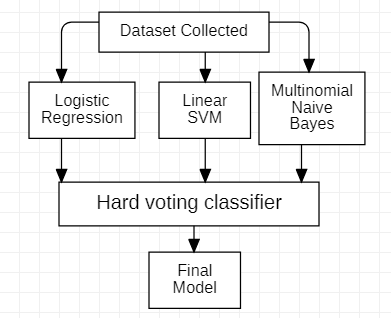


Fig1. Final Model

Conclusion

Through our study we came to know that bi-grams performed better than uni-gram but increased model complexity. Hard voting method using Random forest, XG boost and linear SVM with bi-grams gave a higher accuracy of 85.9% but takes more time to build models and process results due to the huge number of parameters. LSTM and SVM are more practical as they gave accuracy around 84% and took less time compared to voting methods. Sampling of complaints and building a hard voting classifier (*uni-gram*) with linear SVM, NB and Logistic Regression gave acceptable accuracy of 82.4% and processed complaints much faster in real time. Decision trees gave the least accuracy among all models. The choice of the model is based on user need based on speed and accuracy. The future work could be improvising speed while keeping accuracy as high as possible and building fully working software for the finance companies.

Acknowledgement

We would like to express our gratitude to Prof. Ashwini M Joshi, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this Capstone Project also we are grateful to the project coordinators, Prof. Silviya Nancy J, for organizing,managing, and helping with the entire process.

We take this opportunity to thank Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support we have received from the department.

We are deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor, PES University, Dr. Suryaprasad J,Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way.

Finally, this project could not have been completed without the continual support and encouragement we have received from our family and friends.

##### References

1. Mohit Rathore, Dikshanth Gupta, Dinabandhu Bhandari,”Complaint Classification using Word2Vec Model”,*International Journal of Engineering & Technology, 7 (4.5) (2018) 402-404 [1].*
2. Xuesong Tong, Bin Wu, Shuyang Wang, Jinna Lv,”A Complaint Text Classification Model Based on Character-level Convolutional Network”,*978-1-5386-6565- 7/18/$31.00 (f)2018 IEEE [2].*
3. Anditya, Dhamir, Dian, Suyanto, Irfan, Anis, Fedy, Rani*,”*Developing an LSTM based Classification Model of IndiHome Customer Feedbacks”,*2020 International Conference on Data Science and Its Applications [3].*
4. M. Ali Fauzi,”Automatic Complaint Classification System Using Classifier Ensembles”*,Telfor Journal, Vol. 10, No. 2, 2018 [4].*
5. Pavels Goncarovs,”Active Learning SVM Classification Algorithm for Complaints Management Process Automatization”,*978-1-7281-5709-2/19/$31.00 ©2019 IEEE [5].*
6. Chandana C,Neelashree,Nikitha G N,Nisargapriya J,Vishwesh,”Bank Customer Complaints Analysis Using Natural Language Processing and Data Mining”,*International Journal of Progressive Research in Science and Engineering Volume-1, Issue-3, June-2020 [6].*
7. Gutha Jaya Krishna , Vadlamani Ravi, Bheemidi Vikram Reddy, Harshal, Jaiswal,”Sentiment Classification of Indian Banks Customer Complaints”,*978-1- 7281-1895-6/19/$31.00©2019 IEEE [7].*

**.**