



Consumer Complaints Classification Using Machine Learning & Deep Learning

Pramod Kumar Naik ¹, Prashanth T ², S Chandru ², S Jaganath ², Sandesh Balan ²

¹Associate Professor, Department of Computer Science Engineering, Dayananda Sagar University, Bengaluru, Karnataka, India

²Student, Department of Computer Science Engineering, Dayananda Sagar University, Bengaluru, Karnataka, India

Email: sandesh123456780@gmail.com

Article History

Received: 27 February 2023

Accepted: 14 March 2023

Keywords:

Classification;
Segregation;
Natural Language Processing;
Artificial Intelligence;
Machine Learning;
Deep Learning;
Data Augmentation;
Lemmatization;
TF-IDF

Abstract

Complaint handling system used by financial companies are handled by live agents these days, there's a need to move from a system handled by live agents to a system which automatically handles the complaints to increase efficiency & save cost & time. We are planning to develop an automatic financial complaint classification system that automatically deals with the customer complaints by segregating the data & routing it to the right department. We are planning to develop the system by using Natural Language Processing (NLP), Artificial Intelligence (AI), Machine Learning (ML) & Deep Learning (DL) concepts and implement using Python, Jupyter Notebook, etc. The end product will be a web-based application system where customer can register their complaints without having to worry about sending it to right department. (Bejarano) Developed system will automatically segregate the complaints & route it to the right department. 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. (Zhang, Zhao, and Lecun) 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.

1. Introduction

Financial Companies get many complaints regarding their service and they have to handle those complaints quickly and effectively so that the customers who are already angry don't get unhappy and continue to trust the company. The complaints these days are handled by customer support people who can be contacted through mail or through the company's website.

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 tugging themselves due to lack of knowledge about the department in financial companies and there are also chances of them choosing a random department since they are only inter-

ested 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 AI.

This is a real world business problem that is solved by the use of AI and ML. Our project can help many financial companies in the following ways:

- **Faster Response:** Complains are processed instantly and helps the companies to give faster response to customers hence saves time for both customer and bank employees.
- **Workforce Reduction:** No need of dedicated people who are assigned to manually read each complaint and then assign them to respective categories.

Financial Companies perspective:

- Employees in banks come across hundreds of complains daily via mails and social media and it becomes tedious tusk to read long unstructured text and then categorize them as complaints pileup.
- This project will relieve the workload of workers to a great extent.

Users perspective:

- Most of the times, customers may get confused and can forget to specify the actual department of the financial company they are targeting to.
- The project helps users to target the proper department they are complaining to.

Any industry run well if we can keep customers happy, they are the backbone for any industry and therefore their review plays a big role in shaping the companies fortunes. Customer satisfaction can drive the company profits and any decision company takes should be done keeping customers in mind.

2. Literature Survey

In the paper Bank Customer Complaints Analysis Using Natural Language Processing and Data Mining, published in the year 2020, authors Chandana, Neelashree, Nikitha, Nisargapriya, Vishwesh obtained dataset from Bank customer's complaints from Kaggle implemented them using Java. First, an unsupervised method, i.e, LDA (Linear Discriminant Analysis) was used to process the classified texts. Next, t-SNE (t-Distributed Stochastic Neighbour Embedding) was used for data visualisation. Sentence segmentation and it's conversion to Tokenization was also done. We've taken implemented text pre-processing concepts like

Punctuation removal, Stopword removal, Tokenization, Lemmatization, etc, to find the similarity in text complaints regarding their service or product. (Chandana et al.)

In the paper Automatic Complaint Classification System Using Classifier Ensembles, published by the author M Ali Fauzi in the year 2018, 204 records from Sambhat system were used as dataset where, text pre-processing was implemented & where Bag of Words (BoW) were generated. Next, using BoW, training of Machine Learning (ML) models like Naive Bayes, Maximum Entropy, KNN, Random Forest classifiers & SVM was done. For the Naive Bayes classifier; Gaussian, multinomial & Bernoulli NB kernes were used. For SVM linear, polynomial, sigmoid, and RBF kernels were used. In the combination stage, hard and soft voting methods are used. In hard voting, document is assigned to the category which is predicted by major number of classifiers and with the soft voting method, the average of different classifiers are used. (Krishna et al.) As a result, Multinomial Naive Bayes with an accuracy of 80.7% was the best classifier among 5 individual ones. Also, an accuracy of 81.2% was obtained when ensemble method with 3 best classifiers were used. Based on this paper, we've implemented a few machine learning classifiers like Naive Bayes, Linear SVM and Decision Tree, etc. into our model to get the accuracy, precision, recall and f1-score percentages of the models and find the best among them to implement into our proposed web-based application. (M and Fauzi)

In the paper Complaint Classification using Word2VecModel, published in the year 2018, Authors Mohit, Dikshanth, Dinabandhu obtained dataset from Financial Consumer Complaints Dataset, where complaints with maximum word limit of 750 were considered. For data pre-processing, oral complaint was converted to text and tokenization is performed. For embedding layer, they used Word2Vec Model, for which Stopwords weren't removed to preserve more contextual information. (Tong et al.) Then their representations from Word2Vec to GRU (Gated Recurrent Unit) Model were passed sequentially. Next, the output of GRU Layer was passed to MLP with single hidden layer and output layer equal to number of classes. For NLP, training was done using standard back propagation and for GRU, training was done using back

propagation through time where Adam as an optimiser was used. 60:40 split was used for training and testing and model built using Keras library. 4 epochs were chosen as optimal by plotting Loss vs Epochs and Accuracy vs Epochs graphs and 85% classification accuracy. From this paper, we've got to know that they implemented only unidirectional RNN, as for future work, bidirectional and Stacked RNN's might be used as an upgrade to it's predecessor. (Rathore and Gupta)

In the paper, Active Learning SVM Classification Algorithm for Complaints Management Process Automatization, published by Pavels Goncarovs in the year 2019, dataset was obtained from Latvian areuse documents, while Experimentation of comparing Decision Tree with SVM was done for complaints classification task. (Arifianto *et al.*) Text pre-processing was done and data was represented as BoW. Terms with relative frequency greater than 3 was used for further experimentation. Decision Tree algorithm and Sequential Minimal Optimization (SVM) were primarily used where SVM with only 20% data usage performed far better than Decision Tree. Results showed SVM got accuracy of 86% whereas decision tree produced irrelevant results. Based on this paper, we've implemented Linear SVM and Decision Tree classifiers into our model to get the accuracy, precision, recall and f1-score percentages of the models and find the best among them to implement into our proposed web-based application. Eventually, it was found out that SVM was more efficient than Decision Tree even with less usage. (Goncarovs)

3. Scope & Purpose :

The Bank employees, often receive lengthy, unstructured complaints. It takes lot of time and effort by these employees to know the department they need the complaint to be addressed to. (Jiang *et al.*) So this application will organize these complaints without having to actually read the complaints. Also users who are uncertain as to which exact department of the bank they have to complain to can also make use of this application. The objective of this application is to instantly classify complaints, free up manpower. Limitations of this application will be regarding the language the user uses, application would not support all the languages the user uses.

4. Materials and Methods :

4.1. Materials :

4.1.1. Functional Requirements :

- Input : Complaint text will be given as input.
 - Output : Classified results showing departments along with the percentage matching to that respective departments.

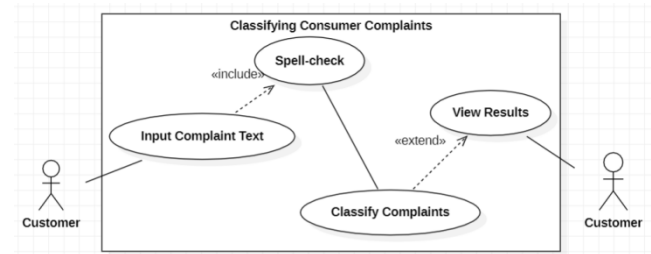


FIGURE 1. Use Case Diagram

4.1.2. Non-functional Requirements :

- Performance Requirements : The performance is mainly dependant on the internal working of machine learning model accuracy. The classification model chosen will provide accurate results while classifying. The application is reliable in terms of taking input and displaying results as soon as possible when users click the submit button. The application will be accessible in any device with internet connection and browser.
 - Safety Requirements : Safety measures have to be taken to make sure that server won't face downtime while serving the web page. Maintenance team is desired for safeguarding the working of the team.
 - Security Requirements : Users tend to share their banking information like account number, credit card/debit card number while complaining. The website need to make sure that there is no man-in-the-middle attacks, so we use https instead of http.

4.1.3. Hardware Requirements :

- Server : The Financial Text Classification application will run on a web server listening on port 80.
 - Client : The web application will be displayed on client's monitor or laptop screen. The application will encourage users to use the mouse to interact with the components of the Website, Mouse will help users to activate buttons like submit and also helps to position the Cursor. The application also needs the requirement of a keyboard if users wish to type the complaint.

4.1.4. Software Requirements :

- Serverside : Flask will be used for backend. It provides a development server and a debugger.
 - Clientside : Any popular web browsers which supports JavaScript, HTML 5, CSS
 - Additional Tools : TensorFlow, scikit learn, etc.

4.2. Methods :

4.2.1. Data Acquisition :

For our analysis, we used dataset from Consumer Financial Protection Bureau government website. All the latest complaints are updated on their website weekly and the dataset is available for open research. We took latest complaints starting from the last year for our analysis. All the complaints description is present in "Consumer_complaint_narrative" column and the category of complaint is present in "Product" column. There are 9 categories in total in the Product column

Since the dataset was highly imbalanced as shown in below figure, we handled class imbalance problem 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 is taken for further analysis.

4.2.2. Data Augmentation using Synonym Replacement:

4.2.3. Exploratory Data Analysis (EDA) :

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

4.2.4. Text Pre-processing :

We applied the following pre-processing steps to our text data:

- Lower Case Conversion
- Punctuations Removal
- Digits Removal
- Stop Words Removal
- Lemmatization
- Removal of confidential information which were represented by x 's in complaints

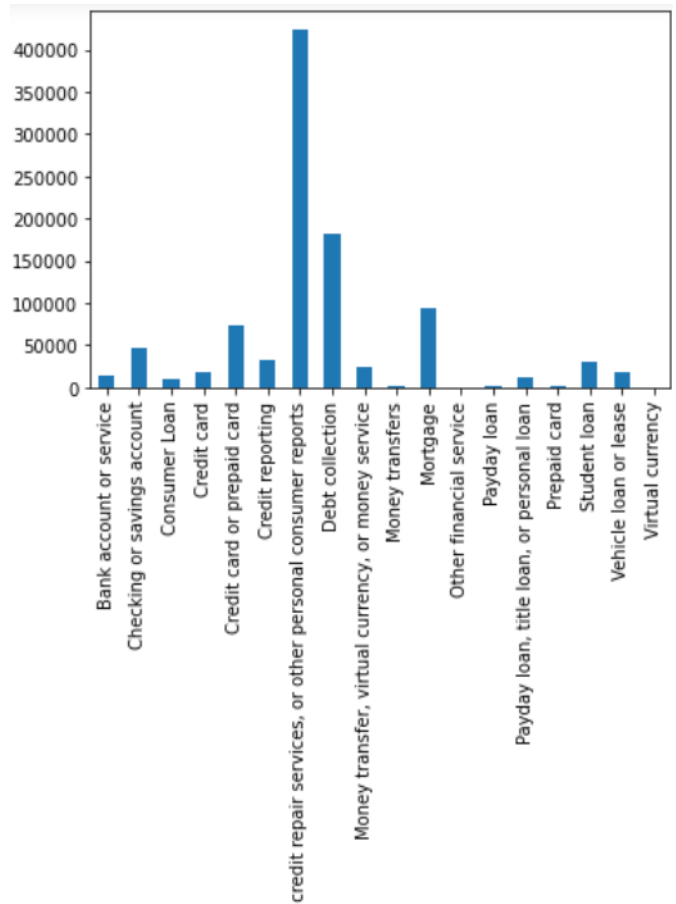


FIGURE 2. : Initial Class Distribution

This is ans: someone in due north Carolina has stolen my identity information and has purchased items including XXXX cellula r telephone speech sound thru XXXX on XXXX/XXXX/2015. A police report was filed as soon as found out about it on XXXX/XXXX/2015. A investigation from XXXX is under mode thru there humbug department and our local police department.

FIGURE 3. : Synonym Replacement

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



FIGURE 4. : WordCloud Vectorization

4.2.5. Machine Learning Model Building & Testing :

For word vector representations, Term Frequency - Inverse Document Frequency technique has been used. Then, the data is split into train and test data of 70% and 30% respectively. 5 Machine Learning algorithms are being implemented and compared.

Multinomial Naive Bayes :

79.82 percent accuracy with Multinomial Naive

Bayes Model with alpha value of one for smoothing has been achieved.

Decision Tree :

Using Gini Impurity Index, Decision Tree Model is built and an accuracy of 71.51 percent has been achieved.

Linear SVM :

Linear SVM Mode, which is inherently based on best fit line that separates the data points, is built. One vs All method is being used to classify the complaints using Linear SVM due to it being a multi-class problem. L2 penalty with squared hinge loss & C value of 1 are used. Using this model, we achieved an accuracy rate of 83.62 percent.

Logistic Regression :

Logistic Regression which is based on logistic function is being built using one vs rest method since it is applicable innately for only binary classification problems. L2 regulariser with a C value of 1 is used. We obtained 83.62 percent accuracy.

K-Nearest Neighbours :

KNN model is the last ML Model that we have built, since it's simple to understand & implement. Since K value is needed specifically for building the KNN Model, Elbow Method technique was used to find the optimal value of K as shown in the figure below. All K values are taken as odd variables to avoid ties. Euclidean distance is used as distance metric.

Using elbow method from the above graph, K value of 11 is taken for building the model and an accuracy of 74.82 percent is obtained.

Machine Learning Algorithms using Bigrams :

After implementing Machine Learning Models using Unigrams, we implemented these models using Bigrams considering 2 words at a time for training. The results obtained are shown in the below table.

TABLE 1. Unigrams vs Bigrams

Models	Unigrams	Bigrams
Multinomial Naive Bayes	79.99%	82.94%
Decision Tree	71.33%	65.35%
Linear SVM	83.51%	85.35%
Logistic Regression	83.03%	81.71%
K-nearest Neighbours	73.76%	69.10%

4.2.6. Deep Learning Model Building & Testing :

Deep learning models in recent times have become popular among researchers for many NLP tasks like machine translation, sentiment analysis, etc. For this phase, a deep neural architecture (LSTM) has been built by us.

Long Short Term Memory (LSTM) :

LSTM is an RNN Model variant which can control long term dependencies effciently in comparison to vanilla RNN and hence it can be very effective for many NLP functions.

Maximum of 400 words from each complaint & embedding dimension of 100 are taken. Along with 10 percent validation split, ratio of 80:20 for Train-test is chosen. The batch size of 64 for number of epochs is set to 9. Between LSTM and input layers, an embedding layer is implemented. 20 percent Spatial Dropout is implemented. Next, LSTM layer with dropout of 128 units and recurrent 20% dropout is used to minimize/avoid overfitting. Final layer consists of a 9 neurons for a Dense Layer since there are 9 classes and since it is a problem of multi-class text classification, softmax activation function is being implemented. Categorical cross entropy is the loss function in addition to Adam optimizer is used.

Patience factor of 2 for Early Stopping is used which implies that the model stops when maximum of 2 epochs shows less to or nil validation set improvement. 0.0001 of Min_delta is set which means to be assessed as an improvement for it in Early Stopping, minimum improvement must be more than 0.0001.

Because of Early Stopping, after 7 epochs, the model stops running. After 7 epochs, Training accuracy , Validation set accuracy & Testing set accuracy percentages of 91.6, 78.48 & 77.86 are obtained respectively.

Word Level Convolutional Neural Networks :

2 Convolutional layers each with 128 filters has been used with kernel size of 5. Relu is the Activation function used in convolutional layers. Max-pooling layer with pool size of 5 along with dropout rate of 30% follows after the Convolutional layers. The rest of the model is similar to LSTM Model & trained accordingly. Testing accuracy is attained at 82.46%.

5. Results and Discussion

This graph & table says about the percentages of all machine learning models used and also gives the user the best algorithm to implement. Firstly for ML models, we came to know that bi-grams performed better than uni-gram but increased model complexity. Decision trees gave the least accuracy (71% for Unigrams & 65% for Bigrams) among all models similar to K-Nearest Neighbours (74% for Unigrams & 69% for Bigrams); in contrast to Multinomial Naive Bayes (80% for Unigrams & 83% for Bigrams), Linear SVM (84% for Unigrams & 85% for Bigrams) and Logistic Regression (83% for Unigrams & 82% for Bigrams) which gave more accuracy in less time. SVM is more practical as it gave accuracy around 84% and took less time compare to voting method. Now, coming to DL algorithms, we inferred that Word Level CNN has more Test Accuracy (82%) & less Test Loss (71%) as opposite to Unidirectional LSTM with Test Accuracy & Loss clocking at 78%. The choice of the model is based on user need based on speed and accuracy.

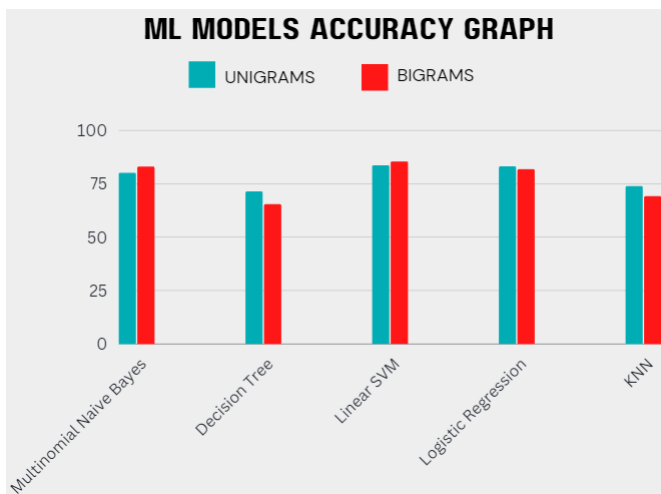


FIGURE 5. ML Models & their accuracy rates

6. Conclusion

We eventually came to know that bi-grams performed better than uni-gram but increased model complexity. Decision trees gave the least accuracy among all models in contrast to Linear SVM & Logistic Regression which gave more accuracy in less time.

It's clear that sentiment analysis and topic classification can be extremely helpful to automatically classify customer complaints, so that you can

respond quickly and efficiently to your customers when they (and your company) need it the most. Putting machine learning and deep learning into practice can help you maintain and monitor your customer base, 24/7, and in real time, finding customer complaints from all over the web and automatically route them to the proper employee.

Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

References

- Arifianto, Anditya, et al. "Developing an LSTM-based Classification Model of IndiHome Customer Feedbacks". *2020 International Conference on Data Science and Its Applications (ICoDSA)* (2020). [10.1109/icodsa50139.2020.9212863](https://doi.org/10.1109/icodsa50139.2020.9212863).
- Bejarano, Gissella. "PhD Forum: Deep Learning and Probabilistic Models Applied to Sequential Data". *2018 IEEE International Conference on Smart Computing (SMARTCOMP)* (2018).
- Chandana, C, et al. "Bank Customer Complaints Analysis Using Natural Language Processing and Data Mining". (2020): 2020–2020.
- Goncarovs, Pavels. "Active Learning SVM Classification Algorithm for Complaints Management Process Automatization". *2019 60th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)* (2019). [10.1109/itms47855.2019.8940658](https://doi.org/10.1109/itms47855.2019.8940658).
- Jiang, Hui, et al. "Network Intrusion Detection Based on PSO-Xgboost Model". *IEEE Access* 8 (2020): 58392–58401.
- Krishna, Vadlamani Gutha Jaya, et al. "Sentiment Classification of Indian Banks Customer Complaints". *Sentiment Classification of Indian Banks Customer Complaints* (2019). [10.1109/tencon.2019.8929703](https://doi.org/10.1109/tencon.2019.8929703).
- M and Ali Fauzi. "Automatic Complaint Classification System Using Classifier Ensembles". (2018). [10.5937/telfor1802123a](https://doi.org/10.5937/telfor1802123a).
- Rathore, Mohit and Dikshant Gupta. "Complaint Classification using Word2Vec Model". (2018).

[10.14419/ijet.v7i4.5.20192](https://doi.org/10.14419/ijet.v7i4.5.20192); <https://doi.org/10.14419/ijet.v7i4.5.20192>.

Tong, Xuesong, et al. “A Complaint Text Classification Model Based on Character-Level Convolutional Network”. *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (2018).

Zhang, Xiang, Junbo Zhao, and Yann Lecun. “Character-level Convolutional Networks for Text Classification”. (2015): 2015–2015.



© Promod Kumar Naik et al. 2023 Open Access.

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License

(<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Embargo period: The article has no embargo period.

To cite this Article: , Pramod Kumar Naik, Prashanth T , S Chandru , S Jaganath , and Sandesh Balan . “**Consumer Complaints Classification Using Machine Learning & Deep Learning**.” *International Research Journal on Advanced Science Hub 05.05S* (2023): 116–122. <http://dx.doi.org/10.47392/irjash.2023.S015>