

## **FINANCIAL CUSTOMER COMPLAINT CLASSIFICATION USING NLP**

**Sneha H A**

PG,Student

Dept. of MCA

The Oxford College of Engineering,  
Bommanahalli, Bengaluru- 560068

[snehahmca2025@gmail.com](mailto:snehahmca2025@gmail.com)

**Sowmya J**

Assistant Professor

Dept. of MCA

The Oxford College of Engineering,  
Bommanahalli, Bengaluru- 560068

[sowmyaj@theoxford.edu](mailto:sowmyaj@theoxford.edu)

### **ABSTRACT**

Improving customer satisfaction and regulatory compliance in contemporary banking sector depends on effectively handling and resolving client complaints. Conventional manual complaint analysis techniques are laborious, prone to mistakes, and unable to manage growing amount of unstructured data. This project suggests a system for classifying financial client complaints that automatically analyzes and groups customer complaints using Natural Language Processing (NLP) techniques. To find complaint categories such loan problems, credit card disputes, transaction faults, and service-related difficulties, system makes use of text preprocessing, feature extraction, and machine learning models. Through the use of sophisticated algorithms and sentiment analysis, technology improves categorization accuracy while empowering financial institutions to quickly prioritize and resolve complaints. In addition to increasing operational effectiveness, the suggested strategy offers useful information for decision-making,

which enhances customer relationship management and builds confidence in financial services.

**KEYWORDS:** Document Classification, linearSVC, Customer complaint classification.

### **INTRODUCTION**

A huge amount of client interactions, complaints, and feedback are generated every day through emails, chatbots, service portals, and social media platforms as a result of the digital financial services industry's explosive rise. Improving consumer happiness, regulatory compliance, and general trust in financial institutions all depend on how well these complaints are handled. Manually going over and classifying complaints is ineffective, time-consuming, and prone to mistakes. Natural Language Processing (NLP) offers a potent remedy for this problem by making it possible to automatically analyze and categorize textual material. NLP models may recognize patterns, extract significant features, and precisely categorize financial complaints into predetermined groups such transaction issues, loan problems, service delays, or fraud-related concerns by utilizing cutting-edge machine learning and deep learning approaches.

The goal of this project is to create a strong NLP-based framework that will help financial businesses proactively address problems, shorten response times, and obtain insightful knowledge about the wants and pain points of their customers—all while increasing effectiveness of complaint management.

## **LITERATURE SURVEY**

Because of the growing number of consumer contacts on digital platforms, customer complaint analysis has emerged as a crucial field of study in the financial industry. Manual review and keyword-based methods used in traditional approaches were laborious and less successful at capturing the semantic meaning of complaints. Many machine learning and deep learning models, including Naïve Bayes, Support Vector Machines (SVM), Random Forest, and Logistic Regression, have been widely employed for text categorization tasks as a result of advances in Natural Language Processing (NLP). Convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), and transformer-based models like BERT are examples of deep learning architectures that have recently shown exceptional performance in comprehending context, sentiment, and intent in customer complaint data.

Numerous studies have emphasized the use of text embeddings (Word2Vec, GloVe, and contextual embeddings), topic modeling, and

sentiment analysis to categorize complaints into groups like fraud, billing problems, service discontent, and transaction failures. Additionally, studies show that financial organizations' decision-making and forecast accuracy are improved when NLP and big data analytics are combined.

## **EXISTING WORK**

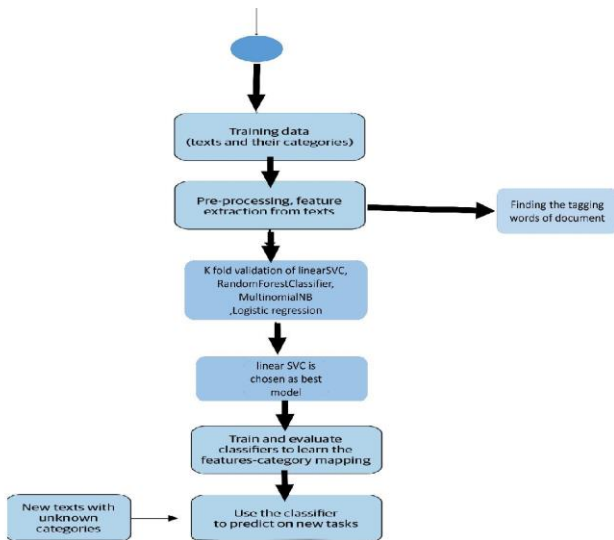
The use of Natural Language Processing (NLP) techniques for the analysis and categorization of consumer complaints in financial industry has been the subject of numerous studies in recent years. Conventional methods used rule-based systems or human tagging, which were frequently laborious and unscalable. Researchers have created automated categorization models that use text mining, sentiment analysis, and topic modeling to better comprehend complaint narratives as a result of the development of machine learning and deep learning techniques. While classifiers like Support Vector Machines (SVM), Random Forests, Logistic Regression, and more recently, deep learning models like Convolutional Neural Networks (CNNs) have been used extensively to represent textual data, techniques like Bag-of-Words (BoW), TF-IDF, and word embeddings (Word2Vec, GloVe, and BERT) have also been used extensively.

## **PROPOSED SYSTEM**

The goal of the proposed system is to create an automated framework that uses machine learning and natural language processing (NLP) techniques to accurately classify financial client complaints into preset categories. Financial clients' complaint data is gathered by system, which then preprocesses it using lemmatization, tokenization, stop-word removal, and text cleaning to guarantee relevant feature extraction. To effectively represent textual data, sophisticated NLP approaches including transformer-based models, word embeddings, and TF-IDF are used. To accomplish classification, these features are subsequently fed into supervised learning algorithms including deep learning models, Random Forests, Support Vector Machines, and Logistic Regression. Suggested solution also includes methods for dealing with unbalanced data and enhances interpretability by offering crucial information about main problems that clients encounter.

## **METHODOLOGY**

The proposed methodology for financial customer complaint categorization automatically classifies unstructured complaint texts into predefined financial product and issue categories using Natural Language Processing (NLP). After raw complaint data is collected, preparation steps like text cleaning, lowercasing, stop-word removal, tokenization, lemmatization, and handling special characters or domain-specific jargon are finished. A term-representation layer is then constructed using advanced embedding techniques like Word2Vec, TF-IDF vectors, or contextual embeddings from BERT/FinBERT in order to capture the syntactic and semantic nuances of financial language. Such feature representations are fed into supervised machine learning classifiers (e.g., Random Forest, SVM, Logistic Regression) and deep learning models (CNN, BiLSTM, or transformer-based architectures) to find discriminative patterns across complaint categories.



Transformer-based architectures, bidirectional LSTMs, random forests, logistic regression, and support vector machines are among classification models that are trained and evaluated. Grid search and cross-validation are used for hyperparameter adjustment in order to optimize model performance. Evaluation metrics include accuracy, precision, recall, F1-score, and macro-averaged AUC, taking into account the potential imbalance in complaint categories. To enhance interpretability and pinpoint the key terms influencing classification decisions, model explainability techniques such as SHAP values and LIME are employed. The best-performing model is eventually deployed as an API endpoint, enabling automated, real-time classification of new complaints, to promote speedier resolution and a better customer experience in the financial industry.

## EXPERIMENTAL RESULTS

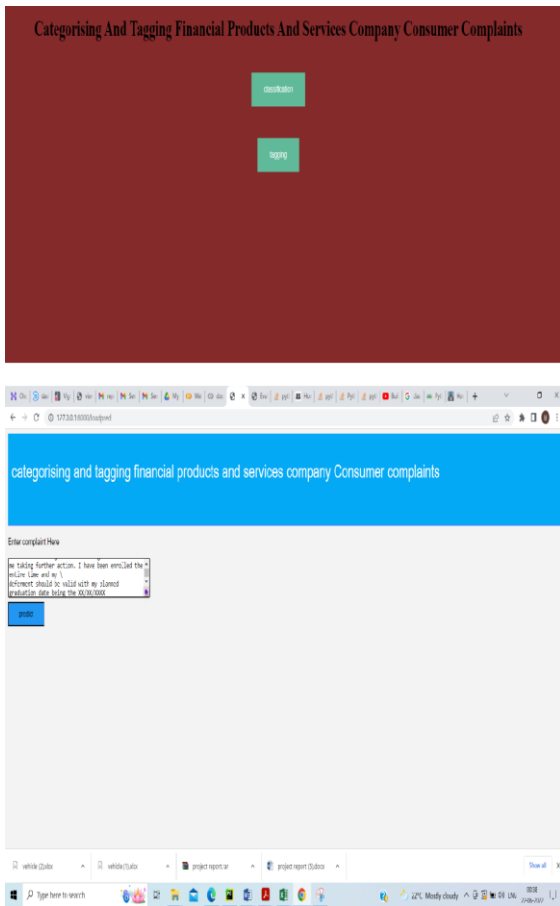
This phase begins after the system has been analyzed, designed, coded, and approved. As of right now, the software is set up to facilitate the desired business operations. System's performance is evaluated using productivity targets established during the strategic planning phase. In addition to installing software on end-user computers used for production and integrating it into regular business operations, users also need to be educated and trained. This phase continues until the system is operating in production and meeting the needs of the designated users.

- This below figure represent data set of project

Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company State	ZIP code	Consumer Tags	Consumer consent provided?	Submitted via	Date sent to company	Company response to consumer	Timely response?	Consumer complaint disputed?	Complaint ID
16/10/2019	Checking or savings account	Checking account	Managing or account	Problem using a debit or ATM card	NaN	NaN	FEDERAL CREDIT UNION	FL 32804	Older American	NaN	Web	05/10/2019	In progress	Yes	NaN	328275
16/10/2019	Checking or savings account	Other banking product or service	Managing or account	Deposits and withdrawals	NaN	NaN	EMPLOYEES CREDIT UNION	WA 98204	NaN	NaN	Referral	05/10/2019	Closed with explanation	Yes	NaN	328228

Customers' textual complaints about a range of financial services and products, including bank accounts, mortgages, credit cards, and loans, make up the dataset.

- This below figure represent of result of user search tab in project.

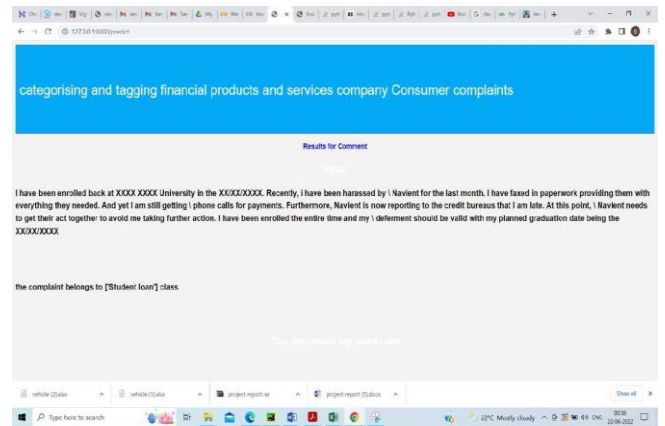


The result shown in the project interface's user search tab is depicted in the figure below. When a consumer types a question or grievance into the search bar, the system uses Natural Language Processing (NLP) techniques to process the input and categorize it into pre-established groups, such as transaction errors, credit card disputes, loan issues, or account management issues.

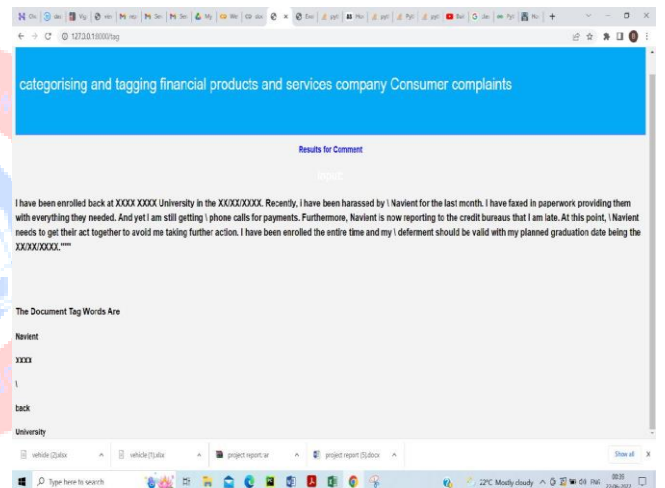
- This below figure represent of result of Result for search in project.

The outcome of search functionality used in the project for classification of financial

customer complaints using natural language processing is shown in the figure below.



- This below figure represent of result of keywords search result in project.



## CONCLUSION

Natural language processing has shown promise in improving the efficiency and precision of resolving customer complaints in the financial industry. System can identify, classify, and prioritize complaints in real time by utilizing machine learning models, feature extraction, and sophisticated text preprocessing techniques. In addition to facilitating quicker problem solving, this aids financial institutions in identifying reoccurring trends that can direct policy and service decisions. The suggested method minimizes response times, cuts down on manual labor, and raises customer satisfaction. Overall, the project demonstrates how incorporating NLP-driven classification frameworks into customer service processes can result in a complaint management system that is more customer-focused, scalable, and intelligent.

## REFERENCES

1. Huang, X., Liu, T., Wang, S., Fu, J., Chen, L., Wei, Z., Liu, Y., ... Huang, X. (2021). Identification of Fine-Grained Elements in Internet Fraud Complaint Text. arXiv.
2. Gao, P., Yang, C., Wang, X., Sun, N., and Zitakis, R. (2023). NLP-based identification of consistent irregularities in customer complaint narratives. arXiv.
3. D. Araci (2019). FinBERT: Pre-trained Language Models for Financial Sentiment Analysis. arXiv.
4. Qurat-ul-ain and associates (2022). A systematic review of computational techniques for information processing from natural language complaint processes. MDPI.
5. F. Vairetti and associates (2024). Consider Before You Classify: The Development of Reasoning Large Language Models for the Identification and Categorization of Consumer Complaints. MDPI.
6. A comparative analysis of fine-tuned deep learning models for financial sentiment analysis and classification. (2024). MDPI.