Enhancing Customer Complaint Classification in Banking: A Deep Learning and Natural Language Processing Approach

MSc Research Project
Data Analytics

Pratheep Kumar Baskaran
Student ID: 22142878

School of Computing
National College of Ireland

Supervisor: Prof. Aaloka Anant

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Pratheep Kumar Baskaran
Student ID:	22142878
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Prof. Aaloka Anant
Submission Due Date:	31st January 2024
Project Title:	Enhancing Customer Complaint Classification in Banking: A Deep Learning and Natural Language Processing Approach
Word Count:	5951
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Customer Complaint Classification in Banking: A Deep Learning and Natural Language Processing Approach

Pratheep Kumar Baskaran
22142878

Abstract

This research pursued a new approach to classify customer feedback in banks using a blend of language processing methods and deep learning technologies. Approach involved a thorough preparation of the data, including converting text to lowercase, reducing words to their base forms, segmenting sentences into words, and omitting common yet non-critical words. The data was also customized to correspond with specific banking products using unique methods. In this study, a key part was experimenting with different high-level models A significant part of the study involved exploring various advanced models. I worked with several types, including CNN, LSTM, and BI-LSTM. These models were applied to two kinds of data: one set that underwent extensive cleaning and another that was processed similarly, but without aligning it with keywords. The findings revealed that the BI-LSTM model excelled, achieving 85% accuracy with keyword matching and 80% without it. Following these insights, a user-friendly interface was crafted for streamlined complaint classification, significantly improving how banks handle and respond to customer complaints. This enhancement in operational efficiency underscores the BI-LSTM model's capability in the detailed task of complaint classification and the broader potential of deep learning in refining essential banking processes.

1 Introduction

1.1 Background and Motivation

When it comes to banking, good customer service is key for building and keeping trust. Nowadays, with most banking done online, how banks deal with complaints is super important. These complaints come in all shapes and sizes, from problems with transactions to worries about digital security. Every single one of these complaints tells banks something about what their customers are looking for and what they value. But dealing with the large number of these complaints is a challenge. Traditional, manual methods of sorting and responding to these complaints are often too slow and not accurate enough for today's fast-paced, data-heavy banking world.

This situation calls for newer, more efficient ways to manage complaints, techniques like Natural Language Processing and deep learning are proving to be very effective. They help sort and make sense of complaints faster and smarter, turning a tough task into something much more manageable. This isn't just good for dealing with complaints; it also helps banks get better at what they do and strengthens their relationships with customers.

In 2022, the recent data from the CFPB, where I sourced the data, highlights some big challenges that they handled around 1.29 million complaints. About 64% (819,800) of these were sent to companies for response. An additional 5% were forwarded to other agencies, and 31% were deemed non-actionable. As of early 2023, less than 0.2% of these complaints were pending. The CFPB expects companies to comprehensively address every issue in a customer's complaint, including providing necessary documentation and detailing any follow-up actions. Companies should accurately categorize their responses, whether offering monetary or non-monetary relief, and aim to respond within 15 days. If a complete response isn't feasible within this period, an interim update and a final response are required within 60 days.

The recent data from the CFPB highlights some big challenges banks face, like getting complaints mixed up because customers pick the wrong product category. This mix-up can slow down getting things fixed and make customers unhappy. I've used NLP and deep learning to tackle this problem. My aim is to sort complaints more accurately, which should help get them to the right place faster and improve how banks handle these issues. This research involves comparing the performances of techniques with and without keyword matching, to find which is more effective. In the end, my work is all about making banking better for customers with quick, right-on-target responses and showing how smart tech can solve tricky problems in banking.. In the end, my work is all about making banking better for customers with quick, right-on-target responses and showing how smart tech can solve tricky problems in banking.

Data point: mortgage market activity and trends 3 [1]

1.2 Research question

"How effective is the hybrid approach of keyword matching and deep learning when compared to deep learning alone in classifying customer complaints in the banking domain?".

In my research, I'm exploring if a mix of keyword matching and deep learning can sort out these complaints more accurately compared to using only deep learning Finding out that using both keyword matching and deep learning together is better could really shake things up for banks It means they could handle complaints more smoothly and keep their customers happier.

1.3 Research objective

Apart from introducing methods for precise customer complaint classification, the upcoming objectives are focused on addressing the research question I presented earlier in section 1.2.

- Developing a Hybrid Model: I'm creating a system that blends keyword matching and deep learning to improve how banks classify customer complaints.
- Comparing Classification Methods: My study will contrast this hybrid model with a deep learning-only method to see which better handles bank customer complaints.

[1] `https://www.consumerfinance.gov/data-research/research-reports/2020-mortgage-market-activity-and-trends/`

- Creating a User Interface: I'll develop a user-friendly interface based on my findings to help banks manage customer complaints more efficiently.

1.4 Structure of document

Moving forward, the report is laid out like, Section 2 dives into a review of research and methods related to our topic. Section 3 outlines the adoption of the CRISP-DM methodology in this study. Section 4 details the research workflow. The implementation is discussed in Section 5. where, Section 6 covers the results achieved. The report concludes with Section 7, where conclusions are drawn, and future research paths are discussed.

2 Related Work

2.1 The Importance of Customer Complaint Management in Banking

Pio et al. (2023) 2023 research offers valuable insights for our study. It investigates customer complaint handling in both traditional and online banks, identifying prevalent issues such as unexpected charges and inadequate service. This research emphasizes the critical need for banks to manage complaints effectively, particularly as banking increasingly moves online. Their work highlights the vital importance of customer complaint management in the banking sector. Hiqmah (2021) 2021 research illuminates the current trends in customer service within the banking sector, especially highlighting the role of social media platforms like Twitter in managing customer complaints in Indonesia. This study is particularly insightful in revealing how effectively addressing complaints online, especially those related to service issues, is critical for ensuring customer satisfaction. It shows the increasing importance of digital platforms in the banking industry, transforming traditional customer-bank interactions. These findings are highly relevant to our research, emphasizing the need for banks to skillfully handle digital complaints, an essential facet of customer service in today's digitally connected world. This study not only provides a closer look at the digital evolution of customer service but also underscores its significance in maintaining and enhancing customer relationships in the banking sector.

The study by Oru and Madumere (2022) explores the influence of customer complaint management on the marketing performance of banks. The research focuses on understanding how banks' handling of customer complaints affects their marketing success. Key findings suggest that effective complaint management can significantly enhance a bank's marketing performance, impacting customer loyalty and satisfaction. This study is relevant to your research as it connects the dots between efficient complaint handling and the broader marketing achievements of banks, reinforcing the importance of effective complaint classification systems in improving customer relations and business success in the banking sector.

2.2 NLP and Machine learning in classification systems

The study Filgueiras et al. (2019) explores case of applying machine learning and NLP in complaint classification, in economic and food safety domains. Their methodology, employing SVMs and LSTM models to address classification challenges like target economic activity and infraction severity, achieved an accuracy surpassing 70%. This study not only highlights the versatility of NLP and deep learning methods in various domains but also presents a compelling case for their potential utility in banking complaint classification, a central focus of my research.

Alamsyah et al. (2022) research presents an innovative method for classifying banking customer complaints using TF-IDF for data preprocessing and Neural Networks for categorization. This method helped Bank Rakyat Indonesia classify concerns about debit and credit cards, customer service, and mobile banking. The study's complaint categorization precision shows usefulness of TF-IDF and Neural Networks, setting a bar for my research. Their preprocessing and systematic classification methods could be changed to enhance banking complaint handling, aligning with study aims. Vinayak and C. (2023) used deep neural networks & word embedding models to categorize customer complaints. This study employed Word2Vec, FastText, BERT, and DistilBERT. LSTM, Bi-LSTM, GRU, and 1D CNN neural network models deciphered language's complexities. Their research efforts are effective, as DistilBERT & CNN received 93% F-score. This technique might increase banks customer complaints classification accuracy and efficiency, which is my study goal. Prabhu et al. (2023) developed a revolutionary complaint automation system. Their system, which uses ERNIE for complex language understanding, excels at text preparation and keyword extraction, improving complaint classification. This method improves handling efficiency and response speed. The installation and success of this approach in critical domains like cyber grievances give significant views for my banking sector research on using advanced NLP methods to enhance complaint management.

Thomas (2018) created a novel technique that utilizes LSTM networks to categorize customer complaints in online forums. Application efficiently classifies diverse questions & complaints, successfully overcoming obstacle of human categorization. Research highlights effectiveness of LSTM network in achieving automation, with an impressive accuracy of 62.825% into inner router & 90.67% in outer router, as determined by validation data testing. This research is very relevant to my research regarding the classification of banking complaints, since it offers a concrete demonstration of utilizing LSTM networks for automated categorization of complaints. Cited accuracies & methodology provide useful insights in possible use of comparable LSTM based technologies in banking sector, particularly for improving effectiveness & precision in managing client feedback. Study conducted by Setiawan et al. (2023) primarily examines use of BERT model in conjunction with many machine learning techniques, including SVM, KNN, Random Forest, & Decision Trees, for purpose of multi-label categorization of student feedback data. Most notable discovery was outstanding efficacy of SVM technique with a linear kernel into multi-label categorization of student feedback. It achieved an amazing accuracy rate of 82% & F1 score of 90%. Use of BERT with conventional machine learning models into multi-label classification offers significant insights for my study on classifying banking complaints. Book emphasizes efficacy of integrating pre-trained language models using machine learning approaches to improve classification accuracy. This method might improve banking complaints categorization. Oyewola et al. (2023) optimized customer complaint management using TSR1DCNN model. A succession of 1D convolutional layers and residual connections enhances this model's complaint data pattern detection. The model had 78.07% - 76.53% accuracy upon training & testing sets. Banking sector applications of advanced neural network models in processing large complaint databases are possible. This may increase complaint classification accuracy & efficacy.

Arslan and Cruz (2023) classified business news using NLP entity recognition and theme modeling. This approach of organizing news texts may increase banking customer feedback accuracy and efficacy. While study doesn't provide accuracy figures, its thorough technique may improve text classification in commercial fields like banking.

This study uses a novel hybrid NLP-machine learning strategy to categorize banking customer complaints. CNN, LSTM, & BI-LSTM's ability to handle complex data patterns & content categories influenced this work's methodology. This decision was inspired by the successful applications of these models in research of Pio et al. (2023) and Arslan and Cruz (2023), displayed their efficiency in processing complex information. These models' capacity to process and analyze large volumes of text data makes them ideal for the diverse and nuanced nature of banking complaints. The hybrid approach is designed to harness these strengths, aiming to en- hance accuracy in classification and reduce misrouting, a key issue identified in previous research This research, therefore, not only contributes to the theoretical understanding of applying NLP and deep learning in banking but also promises practical benefits in im- proving customer service and operational workflows. This approach not only overcomes some of the shortcomings found in earlier methods but also aims to make things run smoother and more efficiently in banks.

3 Methodology

To steer my research, taking cues from the valuable insights shared in the research paper 'Applying the CRISP-DM data mining process in the financial services industry' Plotnikova et al. (2022), I adopt the methodology of choice: the Cross Industry Standard Process for Data Mining (CRISP-DM). This method unfolds in six sequential stages, thoughtfully displayed Figure 1. The intricate details of each stage find thorough exploration in the subsequent subsections.

3.1 Business understanding

In this phase, the groundwork is established by comprehending the goals and necessary achievements. The primary focus is on comprehending customer complaints in banking to enhance feedback management. The emphasis lies in saving time and enhancing efficiency. Through a strategic combination of techniques, the aim is to accurately categorize complaints, ensuring prompt redirection to the appropriate teams. This phase sets the foundation for the entire project, maintaining clarity and a sharp focus on the main objectives.

3.2 Data understanding

The data used in this research comes from the Consumer Financial Protection Bureau and includes consumer complaints against financial institutions. Initially consisting of 903,983 entries spread across 18 columns, the dataset underwent refinement to facilitate a more targeted analysis. Two pivotal columns, namely "Product" and "Consumer Complaint Narrative," were identified to guide the subsequent investigation. A deliberate decision to discard extraneous columns aimed at sharpening the dataset's relevance and efficiency. Consequently, the refined dataset now comprises 903,983 rows and 2 columns. The "Product" column delineates categories such as Debt Collection, Mortgage, Credit Reporting, Credit Card, Bank Account or Service, Student Loan, Consumer Loan, Credit

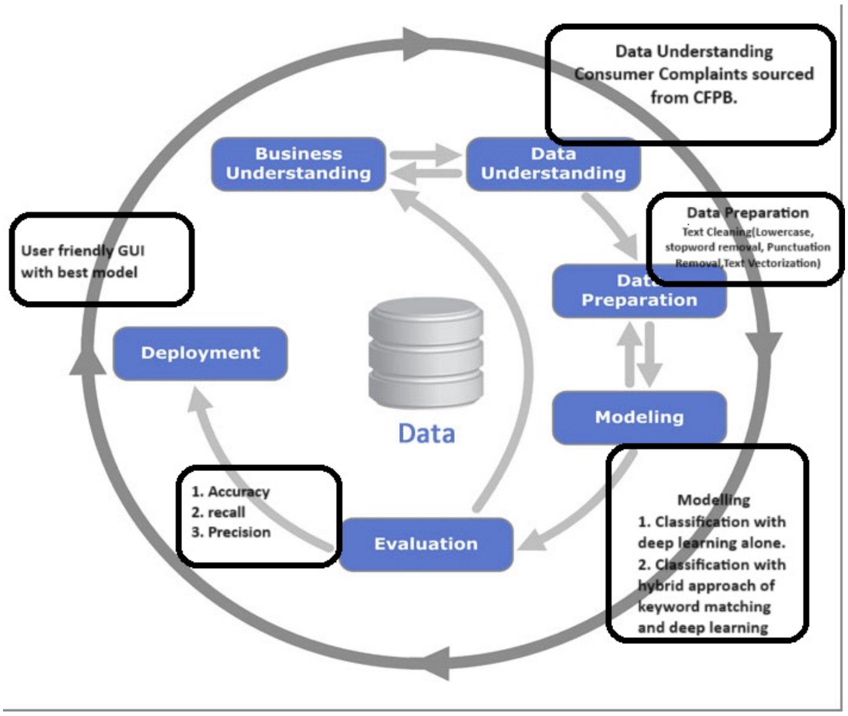

Figure 1: CRISP-DM Methodology

Card or Prepaid Card, Checking or Savings Account, Money Transfers, Prepaid Card, Payday Loan, and Other Financial Services. Simultaneously, the "Consumer Complaint Narrative" column captures the text of complaints submitted by consumers, providing a comprehensive perspective tailored to each financial product category. Data source 3 [2]

3.3 Data pre-processing

Initially, the dataset underwent a refinement process by selecting only essential columns, retaining "Consumer Complaint Narrative" and "Product," while eliminating unnecessary columns. The dataset presented a significant number of null values in the narrative column, primarily because the submission process allows customers to opt-out of sharing their complaint narratives publicly. To maintain the integrity and relevance of my analysis, which relied heavily on textual data, I chose to remove these entries with null values. This strategic reduction aimed at focusing the analysis on key elements. In the process, null values were meticulously handled to ensure data integrity. Specifically, null values in the "Consumer Complaint Narrative" were identified, amounting to 704,013 instances. These null values were subsequently dropped, resulting in a refined dataset with the shape of 199,970 rows and 2 columns.

The data preprocessing pipeline included several crucial steps:

- Transformation to Lowercase: Ensuring uniformity by converting reviews into lowercase. Punctuation Removal: Extracting meaningful information by eliminating unnecessary punctuation.

[2] https://towardsdatascience.com/crisp-dm-methodology-for-your-first-data-science-project-769f35e0

- Frequent Word Elimination: Enhancing precision by removing frequently occurring words.
- Spelling Correction: Leveraging the Text Blob library for accurate spelling correction. Tokenization: Structuring data representation for analysis.
- Lemmatization: Instead of stemming, employing lemmatization for streamlining word variations.
- Classes with fewer than 1000 records, such as 'Virtual currency,' 'Other financial service,' 'Money transfer, virtual currency, or money service,' 'Payday loan, title loan, or personal loan,' 'Vehicle loan or lease,' and 'Credit reporting, credit repair services, or other personal consumer reports were excluded to ensure a robust analysis and meaningful representation of each class in customer complaint classification.

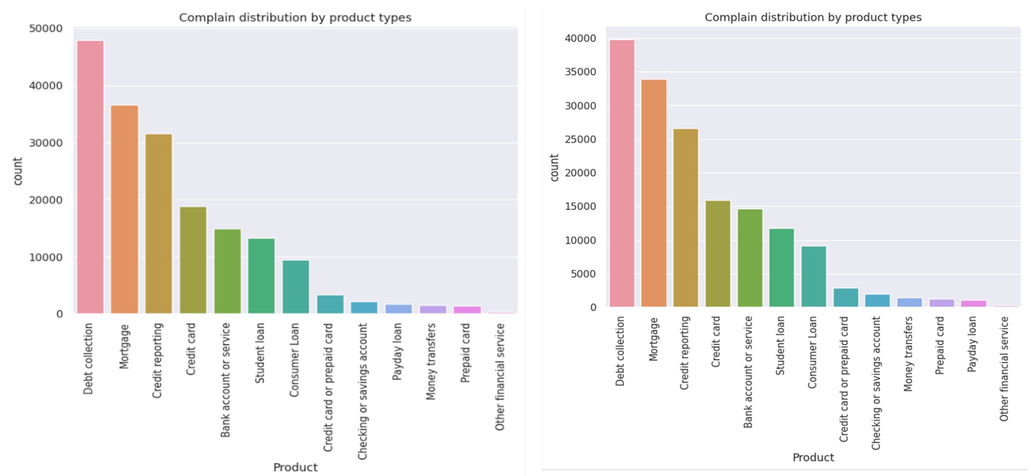

Figure 2: Distribution of products before and after filtering

These actions tidied up our data, getting it all set for the next steps. Think of the tokenizer as a translator turning words into a language the computer gets. This numerical approach helps the computer untangle customer complaints in banking, making our study pack a punch in effectiveness.

3.4 Filtering complaints using Keyword matching

Filtering complaints using keyword matching is a meticulous process that entails crafting and applying carefully chosen keywords for each product category. These keywords are selected based on their frequency within the complaints and their direct relevance to specific domains, ensuring an efficient and targeted preprocessing approach.

Frequency Analysis: Keywords were primarily selected based on their frequency in the complaints. This ensured the inclusion of the most common terms, accurately representing key themes and issues in the dataset.

Domain Relevance: Beyond frequency, I emphasized the relevance of keywords to specific banking domains. This ensured the selected keywords were contextually relevant to banking services and products, enhancing the model's domain-specific effectiveness.

This dual approach of quantitative frequency analysis coupled with qualitative domain relevance assessment ensures the keywords are both statistically significant and contextually appropriate for the banking sector. It's a robust method that significantly boosts the classification model's accuracy. The objective is to categorize and filter complaints effectively, aligning with the nuanced characteristics of each product category in the banking domain.

Product	Keywords
Mortgage	Mortgage, Foreclosure, Interest Rate, Bankruptcy, Fraud, Loan, Escrow, Down Payment, Legal, Trial Period
Credit card	Unauthorized Charges, Dispute, Billing Error, Contract, Cancellation, Chargeback, Terms and Conditions, Fraudulent, Credit, Claim
Credit reporting	Credit Reporting, Score, Credit, Inquiry, Dispute, History, Reporting, Creditworthiness, Freeze, Theft
Debt collection	Debt Collection, Debt, Reinsertion, HIPAA, Legal, Bounced, Credit, Collection, Cease, Desist
Other financial service	Check, Money, Loan, Account, Company, Payment, Bank, Credit, Service, Debt
Consumer Loan	Payment, Loan, Credit, Car, Account, Vehicle, Company, Paid, Late, Report
Money transfers	Transfer, Transaction, Fund, Sent, Bank, PayPal, Western Union,

Figure 3: Pre-defined Keywords for Products

After applying the keyword matching technique to filter complaints, the refined dataset was stored in a CSV format for subsequent processing and analysis. Post-filtering, the shape of the data frame stands at 160,810 rows and 2 columns, providing a focused and tailored dataset for in-depth examination and further stages of the research.

3.5 Modelling

Before diving into the deep learning part, splitted the dataset into training and test sets using cross-validation. The below deep learning models were selected for their own strengths.

1. Convolutional Neural Network (CNN)
2. Long Short-Term Memory network (LSTM)
3. Bidirectional LSTM

Convolutional Neural Network (CNN)- Efficiency in Pattern Identification: Excels in recognizing distinctive patterns or structures within customer complaint narratives, capturing nuanced elements indicative of diverse complaint categories Shahid et al. (2022).

Long Short-Term Memory network (LSTM)- Sequential Language Understanding: Well-suited for comprehending the sequential nature of language, enabling effective grasping of context and dependencies in customer complaint narratives Khataei Maragheh et al. (2022).

Bidirectional LSTM- Enhanced Context Awareness: Brings an additional layer of context awareness, empowering the model to consider the entire narrative context for more accurate complaint classification Khan et al. (2023).

3.6 Evaluation

In evaluating the effectiveness of my research, I utilize a spectrum of standard metrics to assess the model's performance. These include precision percentages, training loss, as well as class-specific recall and precision scores. Of paramount importance is accuracy, which signifies the proportion of accurately identified complaints. As outlined by Sahlaoui et al. (2023), accuracy is determined by dividing the correctly classified instances by the total number of complaints, offering a comprehensive measure of the model's proficiency in categorizing customer grievances within the banking sector. Furthermore, I used metrics like the confusion matrix, specificity, and sensitivity to explore model into the model efficiency in adeptly recognizing and categorizing various complaint types. These metrics provide valuable insights into the model's overall effectiveness, highlighting its distinctive capabilities in handling a multitude of complaint scenarios.

3.7 Deployment

In the practical execution of my categorization project, I conceived a user-friendly interface that simplifies the everyday utilization of the most effective deep learning model. After successfully training and validating the model, I smoothly incorporated it into a user-friendly web interface powered by flask. This meticulous implementation ensures a smooth and intuitive experience for end-users interacting with the categorization system, enhancing its accessibility and usability.

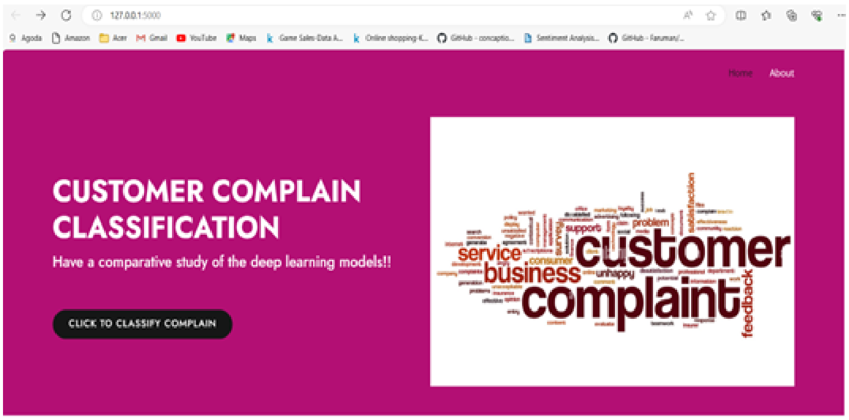

Figure 4: GUI of Web Application

4 Design Specification

During the crucial Design Specification step, expand upon the detailed Data Preparation outlined in the Methodology portion to provide a full plan and execution of customer complaints classification utilizing deep learning models. The provided flowchart serves as a visual guide, leading through the complex process of obtaining data and deploying the model. Provided flowchart acts as a visual guide, leading through complex process of obtaining data and deploying model. This stage include strategic choices, such as data cleansing methods, model selection, & assessment metrics. It acts like the master plan, explaining how we use CNN, LSTM, and Bidirectional LSTMs to make sense of customer complaints. It's like the brain guiding the process of turning words. Consider this Design Specification as a backstage pass, offering a complete view of how I navigate through customer complaints in the banking world. It's like a manual, guiding me step by step, showcasing the decisions I make, the routes I take, and how everything seamlessly comes together to accurately classify those complaints.

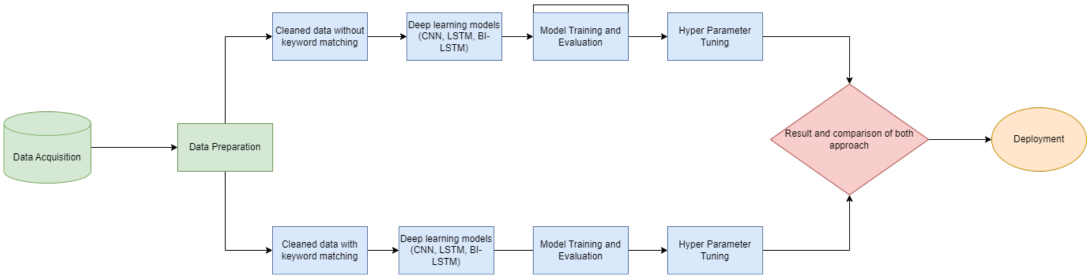

Figure 5: Process Flow Model

5 Implementation

I executed the entire research project using Python because of its flexible libraries for data analysis, visualization, synthetic data generation, and machine learning. The implementation tool of choice was Google Colab, offering a user-friendly web-based environment for creating and executing notebooks. I used TensorFlow with Keras , Scikit-learn, NLTK , Numpy, and Matplotlib for performing this project. The hardware setup included a 12th Gen Intel(R) Core(TM) i5-1240P processor, 16.0 GB RAM, a 64-bit OS, and an RTX 3050 4 GB GPU, ensuring efficient processing for this intensive research. This dynamic combination provided the ideal foundation for seamless and effective project execution.

5.1 Modelling and Hyperparameter Tuning

In this study, I developed three models using keras. The CNN and LSTM models begin with an Conv1D for CNN and LSTM units for LSTM and Dense layers with 'softmax' activation for classification. Both are optimized with SGD and learning rate of 0.01. The BI-LSTM model, with its a Dropout rate of 0.5, and adam optimizer for its adaptive learning, with learning rate of 0.001.

Hyperparameter tuning focused on optimizing learning rates and layer configurations Ali et al. (2023). A preemptive termination method was included in all models to mitigate the issue of overfitting, as stated by Charilaou and Battat (2022). Model's efficacy was assessed using its accuracy upon test data, which ensured that models were not only trained successfully but also verified for their practical usefulness in real-world scenarios. This technique confirmed models' capacity to generalize accurately categorize data in real-life situations.

6 Evaluation

To fulfill study goals, I created classification model utilizing data that had been both cleaned & filtered. Subsequently, I conducted a thorough review of classification report.

Two separate trials were carried out—one using data that had been cleaned, & other using data that had been both cleaned &filtered, with specific keywords being used. The implementation of the keyword matching technique aimed to minimize the risk of misclassified complaints, enhancing the model's proficiency in accurately classifying product-related complaints. Additionally, I conducted hyper parameter tuning to optimize the model performance, a detailed exploration of which will be presented in the subsequent discussion.

6.1 Experiment 1: Modelling on cleaned data without Keyword matching

In the initial phase of my research, I concentrated on analysing the unfiltered, yet cleaned, customer complaint dataset using various models. This pivotal experiment was designed to establish a fundamental understanding of each model's inherent ability to categorize and interpret customer feedback within the banking sector, independent of any specific keyword matching techniques. This approach was instrumental in setting a baseline, providing a clear picture of the models' intrinsic performance in processing and classifying customer complaints in their most natural form.

6.1.1 CNN

The CNN model showed a decent level of accuracy at 66% and exhibited a weighted average precision of 72% and recall of 76%. These metrics indicate a solid foundation in pattern recognition but also suggest a need for refinement in understanding the nuanced context of customer complaints. The accompanying graph of accuracy and validation

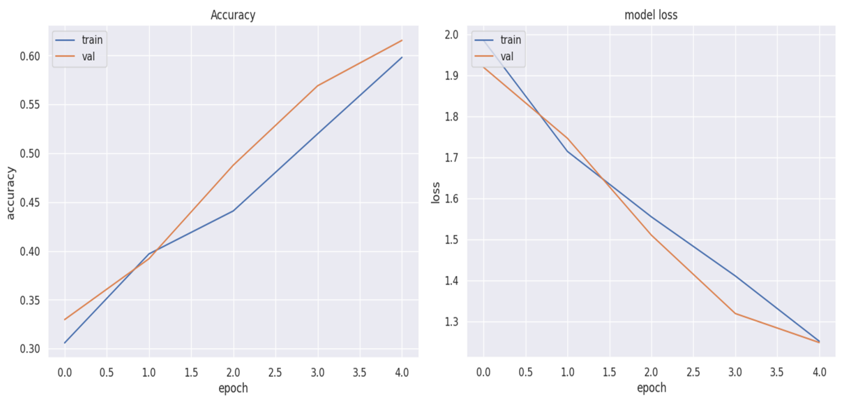

Figure 6: Accuracy and loss plot of CNN

loss illustrates the model's learning curve, indicating areas where model tuning could potentially enhance performance.

6.1.2 LSTM

LSTM model achieved accuracy of 62%, indicating moderate level of comprehension of sequential data. The result indicates that whereas model understands overall structure of complaints, it might be improved by better capturing nuanced changes over time in

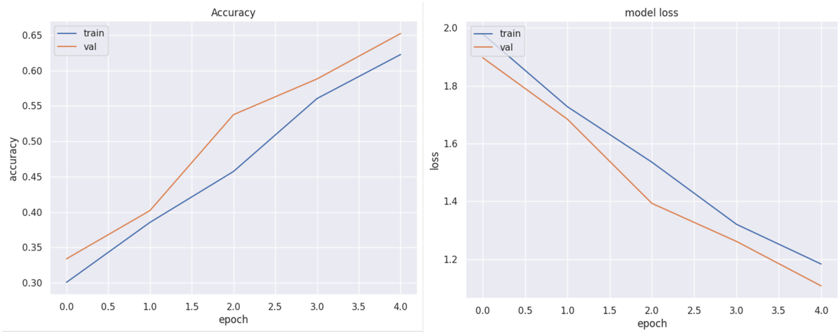

Figure 7: Accuracy and loss plot of LSTM

narratives. Accuracy & validation loss graph for LSTM provides a visual depiction of model's learning progress and identifies areas that may be optimization.

6.1.3 Bi-LSTM

This model performed well, achieving an impressive 80% accuracy.This demonstrates high proficiency in analyzing and understanding content of complaints from both perspectives. Graph illustratin model's accuracy & validation loss unequivocally displays its exceptional performance and speed into learning, underscoring its appropriateness for intricate text categorization tasks such as customer complaint analytics.

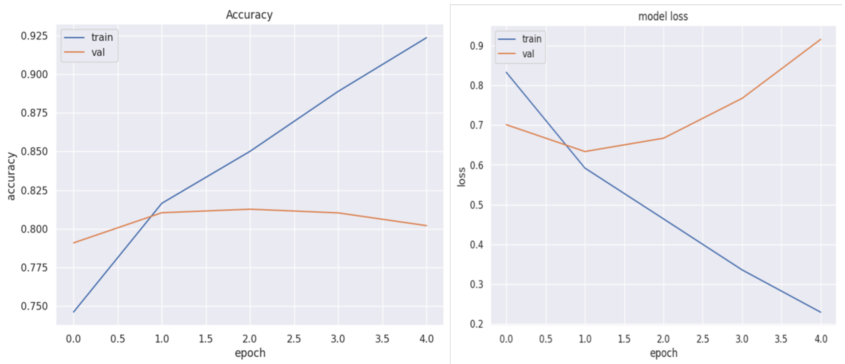

Figure 8: Accuracy and loss plot of Bi-LSTM

Present study demonstrates greater capability of Bi-LSTM model to appropriately categorize intricate customer complaints, hence stressing the advantages of bidirectional computing. Poor efficiency of CNN & LSTM models suggests possibility of improving them via more effective data preparation or model tuning techniques. The training epochs were augmented, and early termination was used to prevent overfitting. Furthermore, distinct learning rates were utilized for every model.

These results are essential for future study, specifically looking at keyword matching. These insights help choose algorithms for specific financial text categorization problems.

6.2 Experiment 2: Modelling on cleaned and filtered data using Keyword matching

This research examines how a hybrid keyword matching method affects deep learning's customer complaint categorization. This process improves cleaned dataset by adding complaint classification-relevant terms. This improves model accuracy and efficacy. Keyword match is tested to see whether it improves models' complaint identification and categorization. This focused method should provide a more contextually rich dataset, helping the algorithms identify each complaint. The current study will help determine practical advantages of match keywords in text processing for machine learning & comparing by original experiment

6.2.1 CNN

CNN model attained accuracy of 71%. This performance is attributed to the increased number of training epochs and the strategic use of keyword matching. The accuracy

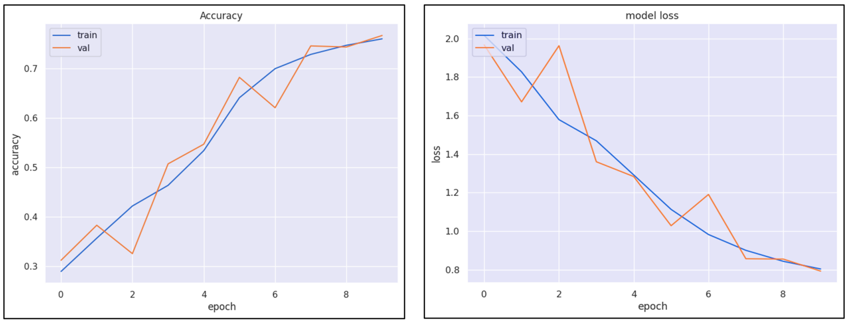

Figure 9: Accuracy and loss plot of CNN with keyword matching

graph for the CNN model exhibits a steady rise over the epochs, indicating consistent improvement

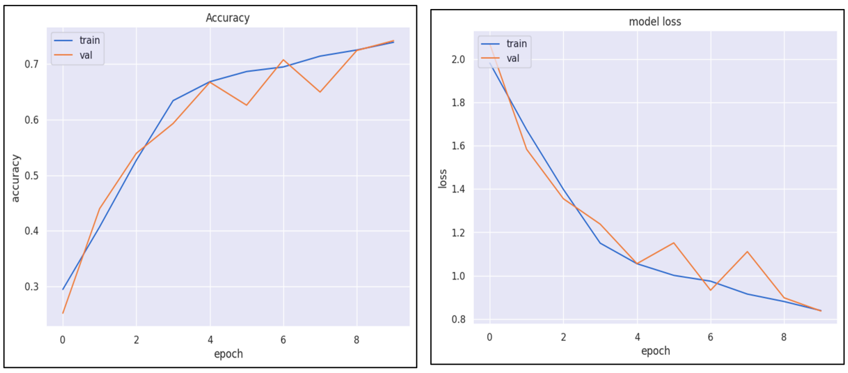

Figure 10: Accuracy and loss plot of LSTM with keyword matching

6.2.2 LSTM

The LSTM model reached an accuracy of 72%. In Fig. . . pattern suggests that the model successfully adapted to the data, benefitting from the extended training period and the implementation of keyword matching, leading to more accurate classification results.

6.2.3 Bi-LSTM

Bi-LSTM model exhibited superior performance, achieving a notable accuracy of 85%. This high accuracy rate is largely attributable to the expanded number of training epochs and the implementation of keyword matching, enhancing the model's contextual understanding of the dataset. The performance metrics, particularly the ascending accuracy graph alongside a descending loss graph, underscore the Bi-LSTM model's efficient learning capabilities and robust performance.

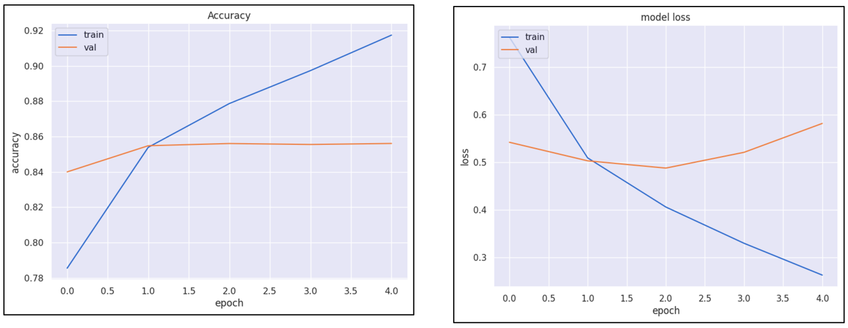

Figure 11: Accuracy and loss plot of Bi-LSTM with keyword matching

Notably, the model's accuracy on the training set showed a continual upward trajectory, in contrast to its validation accuracy, which plateaued at 85%. This indicates a proficient learning from the training dataset, but also highlights a ceiling in the model's ability to generalize this acquired knowledge to novel, unseen data. This observation is pivotal as it illuminates a critical aspect of machine learning models - the balance between learning intricacies from the training data and effectively applying this knowledge to broader, varied datasets.

Algorithm Evaluation Metrics

Algorithm	Accuracy	Precision (Weighted avg)	Recall (Weighted avg)
CNN	71	69	71
LSTM	72	67	72
Bi-LSTM	85	83	85

Figure 12: Evaluation metrics of model with keyword

A noteworthy aspect of Experiment 2 is the reduction in dataset size due to the application of keyword matching. This refined approach, focusing on more relevant data, illustrates a trade-off between dataset size and classification precision. Although dataset was smaller, models demonstrated enhanced accuracy, suggesting efficacy of targeted & quality-oriented data selection for machine learning applications.

6.3 Experiment 3: Comparison of effectiveness of models with keyword matching and without keyword matching

Experiment 3 provides a comprehensive comparison investigation of the efficacy of models that use keyword matching & those that do not for the purpose of categorizing banking customer complaints. Experiment 1 established baseline for model's efficacy by training models on cleaned and complete datasets without using keyword matching. The Bi-LSTM model demonstrated its proficiency in appropriately categorizing consumer comments, although with the limitation of unprocessed data. Experiment 2 used more sophisticated technique by combining keyword matching to clean data, resulting in substantial improvement in accuracy of each model, particularly Bi-LSTM model, that attained an accuracy of 85%.

Confusion matrices of Bi-LSTM model from both techniques are shown in Figure 14. These statistics provide a clear & direct visual understanding of model's accuracy in classifying complaints and the influence of keyword matching upon its efficacy in different complaint classifications.

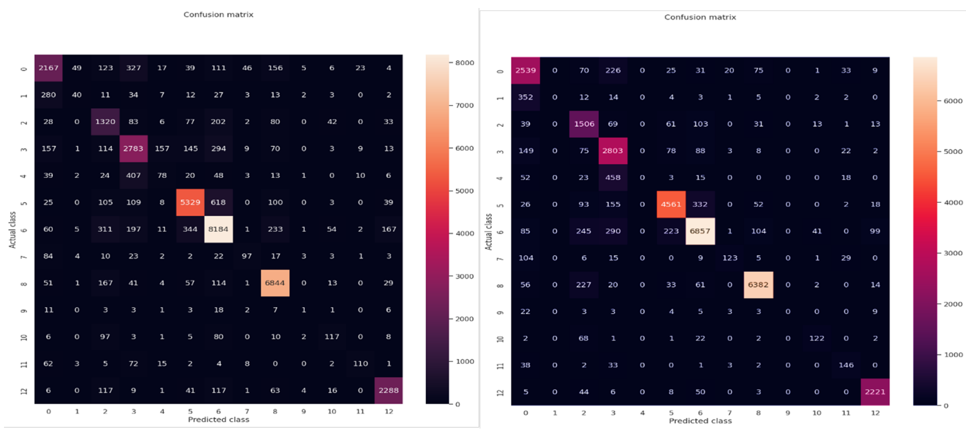

Figure 13: Confusion Matrixes of Bi-LSTM with & without keyword matching

This improvement is attributed to the more focused and contextually rich dataset provided by keyword matching, demonstrating the effectiveness of this hybrid approach.

A crucial observation from Experiment 2 was the reduction in dataset size due to keyword matching, which, despite improving accuracy, prompts questions about the models' capacity to generalize across more extensive and diverse data sets. Notably, the Bi-LSTM model exhibited a plateau in validation accuracy, suggesting a ceiling in its generalization ability.

6.4 Discussion

In my research, I explored the customer complaint classification in the banking sector, emphasizing the use of the Bi-LSTM machine learning model. This research revealed the strengths of a hybrid approach that combines keyword matching with deep learning, significantly enhancing the accuracy of classifying customer complaints. However, a limitation identified in the form of reduced dataset size due to keyword filtering, which impacts the model's ability to generalize. This discovery points to the importance of achieving a balance between the depth and precision of datasets in future studies.

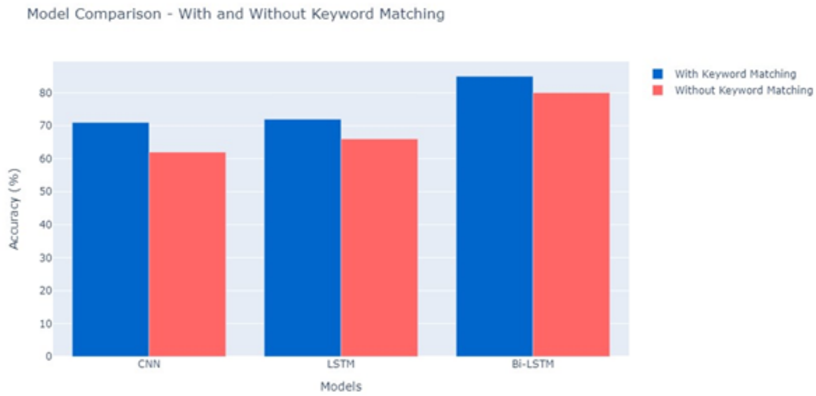

Figure 14: Comparision of Model Performance

The outcomes of my study are in harmony with existing research on natural language processing and machine learning, especially in their use for managing customer complaints. However, my study extends this knowledge by illustrating the specific benefits of combining keyword matching with machine learning techniques.

The successful integration of the top-performing model into a user-friendly graphical interface stands as a significant milestone in my study. This practical application demonstrates the real-world potential of leveraging advanced technology to elevate customer service standards in the banking sector. Future research should aim to explore more diverse and extensive datasets and delve into advanced processing techniques to broaden the applications of these models, potentially including real-time complaint analysis to offer more dynamic solutions in the banking industry.

7 Conclusion and Future Work

In this research, I addressed that: "How effective is the hybrid approach of keyword matching and deep learning when compared to deep learning alone in classifying customer complaints in the banking domain?" The aim was to develop a hybrid model combining keyword matching with deep learning, compare it against a solely deep learning-based approach, and implement a user-friendly interface for the model's practical application. Throughout the study, I rigorously tested various advanced models, such as CNN, LSTM, and notably Bi-LSTM, on two distinct data's, one meticulously cleaned and the other further refined with keyword matching. The most notable outcome of my study was the impressive accuracy of the Bi-LSTM model, which, when augmented with keyword matching, reached up to 85%.

This result underscored the model's effectiveness in clas- sifying customer complaints within the banking sector. This finding not only positively answered my research question but also met the objectives laid out at the study's onset. An important accomplishment of this project was successful integration of most efficient Bi-LSTM model in user-friendly graphical interface. This achievement not only confirmed practical effectiveness of model but also emphasized revolutionary capacity of such technical advancements in enhancing customer service in banking industry. However, research faced several constraints. Reduced size of dataset due to keyword matching, although improving accuracy, raised questions about model's ability to apply to larger & more varied datasets. This discovery emphasizes need for future research to carefully strike balance between amount of data collected & accuracy of models used.

Future study must concentrate upon investigating larger & more diverse datasets, while also using advanced NLP approaches. Furthermore, it could be very beneficial to engage in partnership with professionals from sector in order to find and include additional predetermined keywords. This could enhance model's expertise & precision in certain domains.

Collaborative endeavors of this kind may enhance comprehension of consumer grievances specifically related to banking, resulting in more precise and efficient categorization. In addition, exploring possibilities for commercializing this technology, especially via collaborations with financial institutions, has the potential to transform way banks handle and oversee client feedback, representing a notable progress in industry.

References

Alamsyah, D. P., Arifin, T., Ramdhani, Y., Hidayat, F. A. and Susanti, L. (2022). Classification of customer complaints: Tf-idf approaches, *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, pp. 1–5.

Ali, Y. A., Awwad, E. M., Al-Razgan, M. and Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity, *Processes* **11**(2): 349.

Arslan, M. and Cruz, C. (2023). Leveraging nlp approaches to define and implement text relevance hierarchy framework for business news classification, *Procedia Computer Science* **225**: 317–326. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Sytems (KES 2023).

Charilaou, P. and Battat, R. (2022). Machine learning models and over-fitting considerations, *World Journal of Gastroenterology* **28**(5): 605.

Filgueiras, J., Barbosa, L., Rocha, G., Cardoso, H., Reis, L., Machado, J. and Oliveira, A. (2019). Complaint analysis and classification for economic and food safety, *Proceedings of the Second Workshop on Economics and Natural Language Processing*, pp. 51–60.

Hiqmah, F. (2021). Exploring complaint and complaint management on indonesia's banking industries, *International Journal of Business, Economics and Law* **25**(1): 68–77.

Khan, J., Ahmad, N., Khalid, S., Ali, F. and Lee, Y. (2023). Sentiment and context-aware hybrid dnn with attention for text sentiment classification, *IEEE Access* **11**: 28162–28179.

Khataei Maragheh, H., Gharehchopogh, F. S., Majidzadeh, K. and Sangar, A. B. (2022). A new hybrid based on long short-term memory network with spotted hyena optimization algorithm for multi-label text classification, *Mathematics* **10**(3): 488.

Oru, G. and Madumere, H. (2022). Influence of customer complaint management on marketing performance of banks, *ARRUS Journal of Social Sciences and Humanities* **2**(2): 77–97.

Oyewola, D. O., Omotehinwa, T. O. and Dada, E. G. (2023). Consumer complaints of consumer financial protection bureau via two-stage residual one-dimensional convolutional neural network (tsr1dcnn), *Data and Information Management* **7**(4): 100046.

Pio, P. G., Sigahi, T., Rampasso, I. S., Satolo, E. G., Serafim, M. P., Quelhas, O. L., Leal Filho, W. and Anholon, R. (2023). Complaint management: Comparison between traditional and digital banks and the benefits of using management systems for improvement, *International Journal of Productivity and Performance Management* .

Plotnikova, V., Dumas, M. and Milani, F. P. (2022). Applying the crisp-dm data mining process in the financial services industry: Elicitation of adaptation requirements, *Data Knowledge Engineering* .

Prabhu, A. V., Jefiya, M., Joseph, J. D., Sunny, T. and Abraham, C. M. (2023). Cyber complaint automation system, *2023 Advanced Computing and Communication Technologies for High-Performance Applications (ACCTHPA)*, Ernakulam, India, pp. 1–5.

Sahlaoui, H., Alaoui, E. A. A., Agoujil, S. and Nayyar, A. (2023). An empirical assessment of smote variants techniques and interpretation methods in improving the accuracy and the interpretability of student performance models, *Education and Information Technologies* pp. 1–37.

Setiawan, H., Fatichah, C. and Saikhu, A. (2023). Multilabel classification of student feedback data using bert and machine learning methods, *2023 14th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, pp. 147–152.

Shahid, S. M., Ko, S. and Kwon, S. (2022). Real-time abnormality detection and classification in diesel engine operations with convolutional neural network, *Expert Systems with Applications* **192**: 116233.

Thomas, N. T. (2018). A lstm based tool for consumer complaint classification, *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, India, pp. 2349–2351.

Vinayak, V. and C., J. (2023). Consumer complaints classification using deep learning & word embedding models, *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, pp. 1–5.