Gisma University of Applied Sciences

Gisma University of Applied Sciences

Title: Analyzing Customer Sentiment on Banking social media using Deep Learning

Acknowledgments

I would like to start by stating that I owe a great deal of my present gratitude to my supervisor who provided advice, highly structured feedback, and support that has been invaluable during the process of this study development. Not only led the way to do so but also made me think out of the box and to find clever solutions, with such clear knowledge and experience and under their guidance as a mentor, towards the patients.

I should again like to credit and pay an ode to the faculty and academic staff of my institution, had they provided me with a favorable environment to help me complete this work successfully. They were convinced in scholastic excellence, and they possessed the vehicle in which they conducted the research. I would also like to give credit to my fellow colleagues and peers who provided their keen feedback and discussed and proposed the methodology that got trimmed down.

Among these people I would like to name my friends who were truly trying to be of good faith, who morally supported me and cheered me on when I wanted to give up and who even reminded me that I have to fight. My family too is very happy; much love, patience, sacrifices helped me claim that with all the certainty I possessed that my family was my greatest asset. This faith had given them the strength to keep their faith and stay focused in this process.

Lastly, I would like to give credit to the open-source communities, web resources and dataset providers whose material and tools made actual implementation of this study a reality. The contribution of these people toward knowledge-sharing would have made it much harder to implement this work.

Abstract

This study works out the machine learning and deep learning algorithms to sentiment analysis in banking-related tweets. Major text was cleaned up, preprocessed, and then categorized into positive and negative and neutral sentiments. The outcomes indicated that there are prevailing negative messages of tweets in terms of customer dissatisfaction with the banking services and the word cloud analysis revealed frequent themes that included service quality, transactions, and customer services. Classic models like Naive Bayes and Support Vector Machines, were compared to the methods of deep learning like LSTM, CNN, BERT. Results showed that BERT yielded the best accuracy and F1 scores and that it has better capacity to capture context than traditional baselines. Confusion matrix analysis revealed a regular misclassification problem, especially between the negative and positive tweets, usually caused by sarcasm, slang, or not well-formulated expressions. The research has three research goals such as comprehending the sentimental distribution, comparative analysis between traditional and deep learning, and thematic patterns of customer discourse. The results of the study highlight the importance of explainable AI to financial institutions to get valuable insights on improving the quality of their services, enhance customer relations, and be proactive to new issues.

List of Keywords: Sentiment analysis, Twitter data, Banking sector, Machine learning, Deep learning, Naive Bayes, Support Vector Machine, LSTM, CNN, BERT, Word cloud, Confusion matrix, ROC curve, Explainability, Customer satisfaction

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Background and Context

The advancement in the digital communications space has dramatically changed the level of customer interaction with financial institutions. Social media, such as twitter, has become a very crucial avenue through which the consumers are able to show their level of happiness or unhappiness about the services received by the banking sector, in real-time manner. This shift has contributed to the development of a customer-oriented discourse between banks and the customers where emotion and frustration shared in the form of brief, chatty and not always polite messages can have a significant impact on the industry reputation (Wang et al., 2022). Unlike long-established feedback mechanisms, i.e. survey or call centers, social media can carry a mass opinion dissemination, in real time, that no banks can afford to ignore.

The importance of having an eye on such a sentiment extends beyond the added value to customer experience and to reputational risks management too. A negative tweet can become viral and result in customer loss or can be investigated by regulators. Therefore, banks have taken this road of integrating sentiment analysis tools to assist in harvesting this unstructured data and come up with actionable constraints. The issue with them is that most of these tools are grounded with the old models of sentiment analysis like lexicon-based models or generic machine learning components. In most cases, these models can acquire a rather broad notion of sentiment polarity, but overtones of expression cannot be fulfilled by these models, and it is frequent only directly contextualized notion or a specific direction of action, that can be filled (Dang et al., 2021). It is also a weakness which makes them less competitive in more control-focused and emotionally charged industries such as banking.

The progress made in the sphere of artificial intelligence per se and, namely, in the deep learning segment of these technologies opened the opportunity of possessing the potential to comprehend more complex patterns as far as the language is concerned. Elsewhere we observe other architectures produce modestly more sentiment analysis E2E results, such as Convolutional Neural Networks, Long Short-term Memory, and architecture based on transformers including BERT (Tan et al., 2023). Such models are capable of learning both contextual and semantic features through the massive component of text content accessible that allows an exceptionally careful approach to customer sentiment information description. Their capability to deal with long-term dependence and adapt despite little human interference qualifies them as the perfect prospects to be in the field of studying sentiment in the world of finance where things are continually changing (Asali, 2021).

Given the above tendencies, this study will attempt to add to the existing body of research on the area of artificial intelligence in the field of financial services through building a deep learning model to acquire customer sentiment using analyzing posts left on social media. As the research is narrowed down to the banking-oriented topics, it places itself in line with the general ascent towards the integration of AI into customer relations management, outcome-oriented risk assessment, and personalized financial options (Ray et al., 2020).

1.2 Problem Statement

There is an increasing interest in sentiment analysis in the financial sector, there is still a significant gap when it comes to accurately interpreting usage of nuances and domain-specific expressions present in social media output. With informal language or sarcasm or banking-specific-based allusions, the traditional models draw the wrong conclusions as they are based on keyword-based or hand-built features, which are not reliable discoveries (Asali, 2021). This has been a real issue, especially in an industry where prompt and correct perception of customer mood matters may have an impact on strategic decisions regarding service provision, advertising and crisis control.

In addition, the available sentiment analysis studies have been mostly in the broad areas like product reviews or politics with little consideration to the banking domain. Little research has been done to establish the comparative effectiveness of deep learning architectures in this field, and thus, practitioners have little to guide them when choosing and implementing a model (Dang et al., 2021). Due to the reputational and operational importance at stake, there is a necessity to create more advanced instruments that would allow managing the banking language on social media (Singh et al., 2023). This thus means that this research is timely and has an important need since it involves the application and evaluation of deep learning models in relation to sentiment analysis in relation to banking related to the social platforms.

1.3 Research Aim

To investigate the effectiveness of deep learning techniques in analyzing customer sentiment from banking-related social media content to enhance service delivery and customer engagement.

1.4 Research Objectives

- To develop a deep learning framework capable of accurately classifying customer sentiment in banking-related social media posts.
- To evaluate and compare the performance of different deep learning models (e.g., LSTM, CNN, BERT) for sentiment analysis in this context.
- To analyze sentiment trends and recurring themes within social media conversations about banking services.
- To offer recommendations for banks to utilize sentiment analysis outcomes to enhance customer engagement and improve service delivery.

1.5 Research Questions

- How can deep learning models be trained to identify sentiment polarity (positive, negative, neutral) in banking social media data effectively?
- What is the comparative performance of deep learning models versus traditional machine learning methods in classifying customer sentiment?
- What are the predominant sentiments and themes expressed by banking customers on social media?
- How can insights derive from sentiment analysis inform banks' strategies for customer relationship management and product development?

1.6 Significance of the Study

This research holds significance on academic, technical, practical, and societal levels. From an academic standpoint, it addresses a visible gap in literature by focusing on the comparative effectiveness of deep learning models such as LSTM, CNN, and BERT for sentiment analysis within the specific domain of banking-related social media content. While previous studies have highlighted the general advantages of deep learning for text classification tasks, limited work has been done to evaluate these models in financial or banking contexts where domain-specific language and regulation-sensitive communication prevail (Tan et al., 2023).

Technically, the study benchmarks the performance of multiple deep learning architectures, providing comparative metrics that can guide researchers and practitioners in choosing the appropriate model for sentiment analysis in similar domains (Dang et al., 2021). This kind of performance evaluation, grounded in empirical experimentation, offers valuable insights into model behavior under real-world data conditions.

From a practical perspective, the findings can help banks and financial institutions implement advanced sentiment monitoring systems. These systems could be used to understand customer emotions in real time, enabling improved service delivery, proactive issue resolution, and effective reputation management (Asali, 2021). On a broader societal level, the study promotes the use of data-driven and customer-centric approaches in financial services, fostering trust and responsiveness in customer-bank relationships (Singh et al., 2023). Such advancements not only benefit institutions but also empower customers with better service outcomes (Domingos et al., 2021).

1.7 Scope and Delimitations

This study is limited to the analysis of customer sentiment using textual data from Twitter posts related to banking institutions. The research utilizes only publicly available social media content, thereby eliminating the need for human participant involvement and ensuring full compliance with ethical standards. Posts considered in the dataset are strictly in the English language, which streamlines preprocessing and model training while maintaining focus on widely used banking-related discourse (Wang et al., 2022).

There is also a limitation to text analysis itself without using visual or sound elements like photos, memes, or recordings. Besides, the research fails to examine other platforms other than those that have been offering stable access to the API used and have also been of great importance in banking-related debates which include Twitter. With this kind of scope refinement, the study achieved technical feasibility, consistency in the results quality, and the ethical transparency of the research, without relational deviation to its fundamental purpose of assessing the use of deep learning in sentiment detection to the banking scenario (Dang et al., 2021).

1.8 Structure of the Dissertation

This dissertation presents five chapters with each chapter playing an individual role towards the overall study of how deep learning can transform sentiment analysis of banking-related content on social media.

- Chapter 1 is the background of the study presented, description of research aim, the specific objectives and research questions are stated as well as the establishment of boundaries which are established via scope and delimitations.
- Chapter 2 is a detailed literature review of available literature on the methods used to undertake sentiment analysis, especially in the financial sector and the emerging use of the deep learning models like CNN, LSTM and BERT. It identifies gaps in previous research that this study seeks to address.
- Chapter 3 explains the methodological approach, including the process of collecting and preprocessing Twitter data, followed by the implementation and training of deep learning models.
- Chapter 4 provides an analysis of the results, comparing the performance of the selected models and identifying dominant sentiment trends within the dataset.
- Chapter 5 offers a discussion on the implications of the findings, acknowledge limitations, and provides recommendations for future research and practical implementation in the banking sector.

Chapter 2: Literature Review

2.1 Introduction to Sentiment Analysis in the Banking Sector

Sentiment analysis is the capability of natural language processing (NLP) and machine learning (ML) to be applied to classify unstructured textual data and utilize subjective information in them. When it comes to the use of sentiment analysis in the banking industry, it becomes crucial in deciphering the feedback of the customers which on most occasions is their satisfaction, complaints, or expectations about the banking services. These insights could assist the banks to handle their reputation and enhance the service being offered to their customers (Akter et al., 2024). Sentiment analysis of social media messages, online reviews and customer services interactions, would give banks practical information that has the potential to translate to better experiences and product services to customers.

Sentiment analysis is significant in banking since it helps to evaluate the emotional undertone in the comments of customers, and this is not a fact that can be captured in their detailed analysis. As an illustration, complaints raised by customers regarding the quality of services or concerns raised on products in the bank might be pointing out a greater problem, i.e. the time it takes to respond orris inefficient in the online banking system. It is through analysis of these sentiments that banks can take proactive actions to solve these concerns before they get out of hand. With every increasing level of customer expectations, it becomes vital to take advantage of sentiment analysis as a way for banks to better engage and satisfy the customers (Tan et al., 2022).

Previously, banks tended to incorporate the basic approaches to sentiment analysis, i.e. rule-based lexicon-based sentiment analyzers or classical machine learning. These models, however, tended to have issues with comprehending circumstantial knowledge and became very difficult to comprehend the usage of sarcasm, along with the changing language of customer criticism (Akter et al., 2024). The rise of the deep learning models and especially the Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs) and transformer-based models such as BERT provided a major step forward. These models can provide complex information and infer contextual relations, and this aspect promotes a better sentiment identification (Wu et al., 2024). Today, deep learning techniques are revolutionizing sentiment analysis, allowing banks to gain a deeper understanding of their customers' sentiments and improve their service delivery.

2.2 Traditional Sentiment Analysis Methods

2.2.1 Lexicon-based Methods

The sentiment analysis through lexicon is based on pre-conceptualized list of words that have been associated with certain emotional states, i.e. good or bad. Examples of common typical lexicons are SentiWordNet and AFINN that provide sentiment scores to words according to their tone (Modwey, Abd and Elsamani, 2022). At the same time, these techniques can be easily applied and offer fast outcomes because they match words in the text with sentiment scores. Nevertheless, they are deficient in contextual definition of words. For example, the word "bank" can have a positive connotation when discussing financial services but a negative connotation in the context of environmental concerns. Such context-related meaning may cause wrong sentiment prediction (Onan, 2020).

In addition, lexicon-based systems fail to identify sarcasm, irony, and language nuances that are most often used in customer feedback on social media platforms. The inability to handle these subtleties makes lexicon-based approaches less effective in understanding the true sentiment behind a customer's message, particularly in complex financial conversations (Fattoh et al., 2022). As simple as they may sound, however, all these methods are normally not deemed appropriate in the current sentiment analysis endeavors, particularly where the industry concerned is dynamic, such as the case with the industry of banking, where customer input may be more critical and complex.

2.2.2 Classical Machine Learning Models

Literature Classical machine learning (ML) models have been routinely used to complete the sentiment classification task examples include Naive Bayes, Support Vector Machines (SVM), and Random Forests. Such models are based on feature engineering, whereby several linguistic features applied are the frequency of the words, n-grams, as well as syntactic configurations which are obtained through extraction of the texts, and in turn, they are used as input to the classification (Akter et al., 2024). Although such models perform more adequately than their lexicon-based counterparts in most instances, they still have limitations of features designed manually, which may be tedious, require long time to develop and may not always depict the entire complexity of the language being used in customer feedback.

Moreover, they also exhibit decreased performance with large-scale data, which is characteristic of the banking industry, as it is characterized by millions of customer reviews or posts in social media every day. The models might be inefficient to make general predictions not seen during the training process because of overfitting or long-range dependency problem in the text (Durga and Deepthi Godavarthi, 2023). In banking, feedback can be domain specific and include financial terminology or other regulatory terminology, thus these techniques may be used but need to be customized much more to perform well and therefore become less flexible than modern deep learning models.

2.3 Deep Learning Approaches for Sentiment Analysis

2.3.1 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are the type of neural networks, which process sequential data, say, text, where the current input depends on past inputs in a sequence. Such capability to save prior steps allows RNNs to work especially well on tasks such as sentiment analysis where the meaning of a word or a phrase usually varies based on the context it appears within the full sentence (Modwey, Abd and Elsamani, 2022). RNNs In sentiment analysis, the flow of sentiment through a sentence can be modeled (including determining whether a phrase begins in a positive sentimental direction and ends in a negative one).

The first great benefit of RNNs is that they can identify the relationships across words and phrases across time, which explains why they can be applied to process customer reviews which in most cases can be complex and very context specific. In the review, such as, e.g., “The service was great, but the loan application process took too long,” an RNN can interpret the contrasting nature of sentiments in those sentences by assigning a negative sentiment to the second half of the sentence without dismissing the fact that the service was great earlier in the sentence (Wang et al., 2024). Nevertheless, the RNNs lack the ability to learn long-term dependencies, particularly in text sequences that are long because, as a sequence continues, RNNs easily forget previously acquired information.

2.3.2 Long Short-Term Memory Networks (LSTMs)

Although still based on RNNs, Long Short-Term Memory Networks (LSTMs) are a more elaborate version, and more advanced form of RNNs, and were developed to provide an engine to handle the limitations of a traditional RNN because of the better retention of long-range dependencies. LSTMs have gates (an input, output, and forget gate) which control the passage of information and allow the model to determine what to keep and forget. This results in them being extremely successful at examining lengthy pieces of text where wording / phrases used earlier are essential in determining the overall tone of the text (Tan et al., 2022). As an example, LSTMs can collect and save the sentiment conveyed in the preceding words of a sentence or even two or more sentences in the customer review or social media posts, something that may be lost to RNNs.

The primary application of LSTMs in sentiment analysis is to improve the model's ability to classify sentiments over long pieces of text, such as paragraphs in customer feedback or product reviews. This renders them especially applicable to the study of the affective tone of complex sentences and of even minor changes within sentiment throughout the text (Durga and Deepthi Godavarthi, 2023). LSTMs can therefore serve as effective instruments of analyzing banking sentiment where the insightful customer reviews, such as the mixed ones, need profound contextualization and interpretation.

2.3.3 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have been the norm in image-processing but have also proven to be useful in text-classification tasks. CNNs are created to localize patterns and features within data and thus can be applied to identify the important phrases or keywords which highly exhibit sentiment in customer reviews. For example, CNNs can identify the sentiment-laden words such as "excellent," "disappointed," or "frustrated," and understand how they influence the overall sentiment of the sentence or post (Wu et al., 2024).

CNNs are effective on small scale local spatial features, which is advantageous in using CNN in detecting texts with specific sentiment phrases in text extracts in the short statements of a sentence or the individual review/Twitter post. The CNNs also have a propensity to work even with the existence of noisy labels (e.g. hashtags, emojis) often attached to the posts on social media. Through the acquisition of spatial relationships in text, CNNs have the advantage of processing and categorizing sentiments in different types of banking contents, including complaints, feedback, or praise of products and services (Barik et al., 2023).

2.3.4 Transformer-based Models (e.g., BERT)

Bidirectional Encoder Representations from Transformers (BERT) It is a current deep learning model applying the contextual embeddings to identify the meaning of the word through context. Unlike both RNNs and LSTMs when information is processed at a time, in BERT, a whole sequence of texts can be processed at a time, hence, being highly successful about long-term dependencies and context (Semary et al., 2023). The relationship understanding conveyed by words and the context enables BERT to score higher using previous models in sentiment analysis, particularly in the case of sarcasm or irony or perhaps a complex sentence structure that are ubiquitous in customer reviews or social media posts.

BERT may also be applied to sentiment analysis, which is very useful in the banking industry to determine whether a customer complains about online banking services genuinely or sarcastically. Adjusting the pretrained model with domain-specific data such as banking-related data, BERT has been reported to be shown an immense rise in the accuracy of the classification process as compared to that of the conventional model (Areshey and Mathkour, 2023). This is what makes BERT specifically helpful in comprehending the subtle meaning of the sentiments displayed by the customers of the bank in various social media and other customer feedback platforms.

2.4 Comparative Analysis of Traditional vs. Deep Learning Models

2.4.1 Performance Comparison

Regarding the performance, deep learning models, and particularly those constructed on architectures such as LSTMs, CNNs, and BERT excel over traditional machine learning models such as Naive Bayes, SVM, and Random Forest in terms of sentiment analysis applications. One of the main benefits of the deep learning models compared to traditional ones is that these models can be utilized to learn features directly on raw text without involving any form of feature engineering (Umarani, Julian and Deepa, 2021). It is usually found that deep learning models produce greater accuracy, F1-scores, and precision of sentiment classification, especially with respect to large and complex data as in the case of the banking industry. As an example, researchers found out that

BERT may drastically outperform SVM and Naive Bayes in identifying sentiment nuances (Wu et al., 2024).

Deep learning models have the advantage of learning complicated interactions between words and context, which proves to be very useful in detecting subtle sentiments via customer feedback. In comparison, the traditional machine learning models are likely to fail when handling more demanding structures in sentences or mixed sentiment and involve hand-crafted features, which can be restrictive (Rodrigues et al., 2022). Consequently, deep learning models have emerged as the prima facie candidate in the sentiment analysis application (in businesses like the banking business where customer comments are broader and contextual).

2.4.2 Practical Implications for Banking

There are several advantages related to implementation of deep learning models by banks. This means that first, these models can be more truthful and tamper-free in terms of data on customer sentiment, which leads to increased comprehension of customer concerns, preferences, or pain points (Tan et al., 2022). This will enable the banks to be proactive in involving the customer, hence resolving before it becomes severe, and will enhance customer satisfaction. Also, the role of deep learning models in reputation management includes the identification and diversion of negative reactions about the social media post, which the bank can address using the information and protect its brand reputation in time (Akter et al., 2024).

Besides, sentimental analysis using deep learning can give the banks an informed decision-making process, be it in the optimization of customer service policies, creating new products or improving the current services. Proper sentiment analysis enables banks to adjust the marketing strategies and engage customers with personalized communication to benefit future growth of the business and increased customer retention (Wang et al., 2024).

2.5 Sentiment Analysis in the Banking Industry

2.5.1 Specific Challenges in Banking Sentiment Analysis

The general trends in sentiment analysis on the banking industry reveal that it has a few peculiar issues to face, owing to the nature of language and slang applied by clients and governmental regulated nature of the industry. Data used in context of banking are likely to be filled with terms that are specific to the banking sector, like interest rates, loan processing, credit risk, etc., which might not be comprehensible to a conventional model of sentiment analysis (Modwey, Abd and Elsamani, 2022). Along with that, the financial institutions must also address such complex and sensitive matters and concerns like compliance, privacy and regulations, and so the sentiment analysis process is also not that simple. Making the changes to the phrases on the regulations by mere compliance audit would have resonated 100 times in another environment and hence the models must find the right combination of the words (Akter et al., 2024).

BI dimensionality the other problem is that the industry is its own language with the partners that are in common with the customers themselves. The words such as banking fees, and any payments due, high interest rates etc. are highly emotional based words which is not what would result on simple models. Reviews on social media pages, or in online feedback on the services of banks, contain mixed sentiment, as customers often mention how they feel annoyed or do not understand something about banking services or describe their feelings of satisfaction with them in a way that traditional sentiment analysis algorithms can hardly interpret (Fattoh et al., 2022). This means that we require more advanced models which can interpret such nuances and more so when the information is being generated by a variety of different sources and which are informal such as social media.

2.5.2 Existing Studies on Banking Sentiment Analysis

In the banking field sentiment analysis has been discussed in several studies with the key point of interest in customer reviews and feedback that is received on social media, online forums, and customer service portals. Nevertheless, certain research has also used the classical models of machine learning called Naive Bayes and SVM in sentiment classification, but these models have demonstrated being low efficient in the interpretation of complex and context-specific expressions of customers (Onan, 2020). As an example, a research paper by Wang et al. (2024) used sentiment analysis on financial social media data and discovered that the machine learning models used were unable to classify the sentence with sector specific language and regulatory terms typical of the banking industry.

Furthermore, sentiments classification has also been done utilizing deep learning techniques especially by utilizing RNNs and LSTMs (Tan et al., 2022). These models performed superiorly when it came to taking note of the sequential tendency of sentiment in customer feedback, where longer reviews were more representative. Nonetheless, the literature regarding the specific linguistic features of data related to the banking sector is still lacking, and additional attention should be paid to conduct further research to provide these models with a better performance to make Sentiment classification more accurate and specific.

2.5.3 Trends in Customer Sentiment

Some of the themes of the social media discussions revolving around banking normally include complaints concerning the way customers are being served by the bank, problems with online banking systems, promptness in answering customer questions and openness in the operations of the organization. A major part of feedback involves complaints about slow customer services, problems in accessing the loan processes, or problems of digital banking services (Durga and Deepthi Godavarthi, 2023). These prevalent themes indicate that sentiment analysis could be useful in determining these areas of pain and advise the banks on areas that should be improved.

Sentiment analysis in this example can offer possible practical steps that banks can take to better serve their customers, making processes more efficient and better products. This is possible by analyzing the customers' frequent complaints on the mobile banking app or learning the long queues to borrow funds. Being able to track the customer sentiment in real-time will enable banks to respond rapidly to any negative customer sentiments, limit the damage in terms of reputational risk, and help build positive customer relationships (Deora et al., 2025).

2.6 Research Gap in Banking Sentiment Analysis

2.6.1 Limitations in Existing Literature

Though sentiment analysis has become a significant issue in the banking sector, one of the crucial gaps in literature is the lack of emphasis is given to deep learning models applicable to the bank-specific data in social media. Other studies have used traditional machine learning (SVM, Naive Bayes), which is restricted using feature engineering and the inability to deal with the specifics of the sectoral language (Umarani, Julian and Deepa, 2021). Most of the available models find it difficult to handle the complex vocabulary employed in money-related and financial fields that result in reduced accuracy of sentiment classification.

Moreover, there is a limited selection of studies of models that might encompass the regulatory environment and financial terminology that commonly occur in the feedback of customers concerning the banking industry. Although deep learning methods, such as LSTMs and BERT are successfully used in other industries, their usage in the banking industry, specifically in data processing of social media, is poorly studied (Rodrigues et al., 2022). This leaves the question unanswered in the literature because the models of deep learning would go a long way to make sentiment analysis work effectively given the nature of the language that is specialized by banking customers.

2.6.2 The Importance of Addressing the Gap

It is important that this gap is addressed because it will help advance sentiment analysis within the banking industry. This study will also contribute to the capacity of banks to clearly differentiate and make meaning of customer sentiment through its concentration on deep learning models. An understanding of contexts between words and phrases can be learned deep learning BERT-style models, such as: words in a jargon used in the industry and regulatory jargon the classic model has no knowledge of (Areshey and Mathkour, 2023).

The discovery of such gap in the investigations hitherto made would furnish what might be of more value, something substantive, in other words, in the manner of a piece of customer feedback, something which the banks themselves can apply far more substantially to the explanation of joys and dissatisfactions of the customers in general, and to the later modification of dispensations which they are already making, in general. More than that, the fact of the possibility of providing the means of providing the pleasurably environmental consideration to the customers who render the services of the brightest methods will presuppose the presence of the information that the banks will be capable of being sure of the possibility to introduce the difference in the services they provide as the means of guarantee of the further increase of the number of customers offering the devotion (Wang et al., 2024). Such research one day could yield a complete sentiment analysis model and that in turn would help focus appropriate attention on the problems with which the banking industry has struggled.

2.7 Chapter Summary

The literature review further illuminated a deeper commentary on the topic of sentiment analysis within the banking sector and how the practice has evolved globally and locally in the years preceding the implementation of less developed and more developed deep learning algorithms. It must be said that after manipulating the time-honored instruments of the sentiment analysis i.e. lexicon-based and the time-honored coupled system of machine-learning based methodology in a rather right and proper way, they cannot comprehend the sentiments described by the customer appreciation in a distinctive and delicate fashion, especially the banking department. By contrast, Deep Learning architectures, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), Transformer-based models, including BERT, have demonstrated a better ability to learn contextual dependencies, work with sector-specific jargon and process large-scale data extracted on social media.

One of the research gaps that are to be addressed concerns the inadequate use of deep learning models which are suited to the banking-specific language and regulatory issues. The customer sentiment studies thus far have not exhaustively captured the nuances that are present in this area, especially concerning the merit of analyzing financial terms, including the banking jargon.

Such a research gap can be addressed by the proposed study implementing state-of-the-art deep learning approaches to sentiment analysis that relies on banking-related social media information. The scope of the obtained information during the literature review will also contribute to the methodology development of the study by proposing to utilize deep learning models in understanding sentiment. The observed result is that more accurate customer feedback analysis tools provided to banks will enhance the process of decision-making, reputation maintenance, and service quality improvement.

Chapter 3 Methodology

3.1 Research Design and Philosophical Underpinning

The study aims to be quantitative, experimental, a secondary-data study where the level of sentiment towards banking institutions using social media is established. The ability to analyze huge text data using measurable and replicable input on consumer perception is in line with the general goals of the comparison to the algorithmic performance. According to the basis of positivist philosophy, the design governs the expressions in text as phenomena that can be measured and objectively analyzed. The methodology can be classified as deductive, because it is aimed to validate pre-determined theoretical assumptions regarding the better predicting performance of deep learning models as compared to traditional machine learning classifiers. The specified approach of the study gains further validation through previous research, which demonstrated that social media platforms can be great repositories of data about customer behavior in the field of mobile banking (Asali, 2021).

3.2 Data Sources and Sampling

Twitter/X will be the main source of data used in the study as it is popular, easily accessible, and appropriate to be used in capturing real-time opinions through opinion mining. The collection of relevant tweets will be performed by means of keyword- and hashtag-based searches, paying attention to the names of banks, words related to the services offered, and the events within a certain period. The target of this methodology means that noise is as minimized as possible, and the dataset is still as contextually significant as possible. Additional data cleansing will be done by inclusively considering only public tweets and the tweet will be filtered out based on retweets with no comments, advertisements and irrelevant spam. Also, posts written only in English will be saved to maintain the same level of text processing. PracxdIDB target=returnsDespite the previous studies indicating that careful sampling methods enhance both the quality of sentiment data and the stability of its model results (Bello et al., 2023), subsequent research denotes that less-than-rigorous sampling strategies have taken root in the literature (Friedberg, 2023).

3.3 Data Collection and Preprocessing

The Twitter API will be used to gather data about platform policies and ethics. After extraction, the text corpus shall be subjected to preprocessing to increase its quality and practicability in model training. There will be the removal of duplicate tweets, detecting bots generated data, anonymization of personal identifiers and normalization of text. Emojis and hashtags will be kept, as it is unnecessary to eliminate sentiment-laden elements of language, whereas addressing the domain-specific issues of informal spelling, slang, and financial terms. Earlier tests on text preprocess strategies highlight their strategic position in correcting proficiency, and results stress that regular cleaning and standardization will have a significant impact on the performance of the sentiment model (Palomino & Aider, 2022).

Integration of SMOTE for Balancing

This was with the aim of addressing the problem of class imbalance with the dataset, Synthetic Minority Over-sampling Technique (SMOTE) was applied to develop synthetic samples of the minority group, which contained individual samples balanced in terms of Positive and Neutral sentiments. The idea behind SMOTE is to artificially create new balances of the minority group based on the existing balances to do away with the possibility of bias among training models due to disproportion. It was necessary to enhance the effectiveness of the model to adequately group the Neutral and Positive sentiments which were not happening bearing in mind that the original set of data was skewed to the Negative one.

Using SMOTE, a more balanced model was trained and therefore the model was able to handle the generalization of all the classes of sentiment better. It was timely especially sentiment analysis of financial services when you need the capability to categorize the Neutral sentiment as it should be to listen to the customers.

3.4 Sentiment Labeling Strategy

The data would then be classified under a weak supervision model where a proxy score with positive, negative, or neutral labels is implemented through the presence of sentiment language structures e.g. emojis and hashtags. Although this approach will allow compiling a large, annotated dataset effectively, it also creates a problem of potential noise. To mitigate this weakness, a subset of the data, stratified, will be manually coded, human-validated, with quality controls and label reliability estimation. There will be exclusion of cases that are considered ambiguous to maximize distinction of the classification categories. To demonstrate the fact that the usage of such schemes could allow scaling sentiment analysis and, moreover, to decrease the so-called unnecessary step of time-eating and expensive manual labeling, one could refer to empirically grounded experiments of weak supervision of natural language processing (Lison et al., 2021).

3.5 Model Development

To compare the various approaches to analysis, the modelling approach has been premised on a two-pronged approach to methodology. On one side, we are training classical machine learning baselines, i.e., the Naive Bayes and the Support Vector Machines and training them on the TF-IDF features and superimposing the corresponding predictions on the same baseline according to which we can test predictive performance. The justifications to select the subsequent models are the previous checked capabilities in the process of sentiment classification exercises, designed to execute the consistent outcome at a comparatively small calculating demand. Conversely, deep learning models such as the Long Short- term Memory (LSTM) networks, Convolutional Neural Networks (CNN) and BERTs based on transformers are integrated to extract the context and semantic cues in text. They also include embedding techniques embedding with GloVe or BERT embeddings to improve the quality of a representation. To promote fairness, hyperparameter tuning is performed in a systematic manner and the balance between classes addressed e.g. by weighted loss functions or resampling. The previous comparative research indicates that deep learning techniques tend to perform better in recognizing higher-order sentiment patterns than commonly used classifiers, which is why it is essential that they should be included in this study (Singh & Kaur, 2021).

3.6 Experimental Setup

To ensure rigor in assessment, dataset is partitioned to training, validation and testing groups and time segregation is also made on the dataset to avoid information leaking across the temporal barriers, to minimize the risk of information leaking. The training and evaluation of the models will be approached on CPU and GPU-accelerated infrastructure, where promising libraries like the PyTorch and TensorFlow deep learning frameworks and scikit-learn the traditional baselines to use are already available. The reproducibility is supported by random-seed fix, hyperparameter documentation, and script/configuration version control. Moreover, restrictive measures are imposed to make the comparison between models' fair, e.g., using the same preprocessing pipelines and similar training budgets. Consistent design and replication Sentiment analysis study with some previous experimental work, this review emphasizes the role of systematic setup and reproducibility in sentence analysis studies to ultimately conclude that performance variations can be attributed to models and not variation in procedure (Dhola & Saradva, 2021).

3.7 Evaluation and Analysis

The models are all evaluated in the number of performance measures and their measures are Accuracy, Precision, Recall and F1-score. Macro-averaging is done to bring down the classes' imbalance. Later, to achieve a better conception regarding quality of classification, discrimination capability of model and strength of model in terms of probability that gets estimated is also freed by the use of modified ROC-AUC and calibration. Statistical tests conducted with the purpose of proving the statistically significant differences in the performance are the McNemar test and bootstrap confidence intervals. The error consideration has been performed and the problems associated with sentiment classification such as sarcasm, negation and slang and domain language view are also considered. In addition to the quality of testing of the classification, a trend analysis is also done and it tests the ability of sentiment change and trends over time amongst banks that provide some indication of the dynamics of the consumer sentiment changing environment. Google Scholar provides the list of the current research in which the validity of effective testing strategy and post-visualization research in understanding the current development of deep learning and declared reasonable (Alyoubi and Sharma, 2023).

3.8 Ethical Considerations and Data Governance

The paper has put a lot of emphasis on ethical research to protect the integrity of the research process and protection of the rights of social media users. All the information will be publicly available and all the information that is collected will involve no access to the personal content. Identifiers of the user like profile information and usernames will be either pseudonymized or deleted to preserve individual privacy. Approval mechanisms will be monitored to reduce risks, and it will ensure that the study is tightly within platform terms of service. Safe storage of information will be established, and data will be encrypted to prevent sensitive information and retention periods made clear. Transparency is enhanced through recording methodological choices and being conscious of the possible risks. This was supported by the observation of previous studies on sentiment analysis with the use of the BERT that a robust ethics governance practice must be applied in the utilization of content comprised of user-generated text that social media is too big and too sensitive (Bello et al., 2023).

3.9 Limitations and Validity

The methodology will have some limitations despite the intense design. First, there can be a bias in sampling as Twitter users might not be representative of the entire customers of banks. Second, ineffective labeling can add noise, especially when the labeling concerns sarcasm, irony or linguistic ambiguities. Also, although the models have been optimized to work with English-language data, portability across languages and cultures is weak. These risks are to be appeased through aiding validity through the application of cross-validation, robustness checks, and sensitivity analyses. Time-based testing and subgroup performance checks can be used to consider that the results are not hyper-biased to the factors of time or platform especially. The state of the art in comparison of sentiment analysis appreciation has substantiated that methodological trade-offs, such as labeling strategies and representativeness, cannot be completely avoided but they will be minimized with a proper design and validation (Singh & Kaur, 2021).

3.10 Chapter Summary

In this chapter, I have discussed methodological design, i.e. research philosophy, data sources, preprocessing, labeling, model development, and evaluation measures. Classic machine learning methods and deep learning methods are implemented to evaluate the performance of sentiment classification, and a strict experimental program controls the validity of the comparison. There have also been ethical considerations and limitations because they have been discussed to add transparency and accountability. Overall, the methodology is designed to address the research questions to show how the models can be constructed, measured and interpreted to offer insights into the consumer opinion on the banking institutions. The following chapter will present the findings of such methodological procedures, both on model performance and presenting the patterns which emerged in the data.

Chapter 4 Results And Analysis

4.1 Introduction

In this chapter, the findings of sentiment analysis conducted on the Twitter dataset that was collected will be given, and the purpose of this step is to respond to the research objectives that have been mentioned at the beginning of the work. The analysis aims to reveal the pattern of the opinions concerning banking services, determining the level of traditional and deep learning models, and giving ideas on the dominance of either positive or negative or neutral sentiments. The results are presented in a section with descriptive statistics, model performance contrast, error analysis, and interpretation evaluations using visualization outcomes.

The findings are directly connected with methodology presented in Chapter 3, in which a secondary data method based on Twitter posts has been used in combination with machine and deep learning sentiment classifiers. The chapter does not only explain the results of the models, but also the connection to the strategy of control of consumer attitudes in the financial sphere. In such a way, it incorporates both technical review of algorithms and applied issues in customer experience management in the banking world. Interestingly, the same solutions have been identified in the recent reviews of sentiment analysis studies, where both methodological reliability and usefulness in the real application on the industrial scale are deemed equally critical (Tan et al., 2023).

4.2 Dataset Overview and Sentiment Distribution

The data used in this paper will be 31,962 tweets categorized into negative tweets, positive tweets and neutral tweets. The first examination found that the negative tweets are the majority than other positive and neutral classes. The other half of the customers is annoyed by this contradiction to seeking the services of the other social media that with the help of which the findings of customer attitude investigation will not be familiar now in the banking enterprise (Akter et al., 2024).

To gain a visual representation of the data, a sentiment distribution graph was created that shows an apparent bias towards negative expressions of perception. This alone shows that a displeased client will have an opportunity to share his/her experience with other clients more than a satisfied one. The prevalences of such patterns are relevant since each will suggest that automated models are more likely to be presented with negative than positive examples, which may bias the subsequent classification downstream unless properly addressed.

This imbalance has two implications. One is that model training can develop bias along the lines of the majority class, thus leading to worse predictive accuracy on underrepresented sentiments. Second, in terms of practicality, the proportion of discontent can be misinterpreted by the banks in case they fail to consider the natural behavior of users to criticize more frequently than to reward. Judging by the previous investigations regarding analyzing the data on customer feedback in the banking segment, complaint-related skews in the datasets are a common occurrence that must be considered when constructing sentiment analysis pipelines (Modwey et al., 2022).

This section lays a background to the rest of the sections through the presentation of distribution of sentiments. The imbalanced distribution also indicates the need to utilize heavyweight models and assessment measures capable of dealing with class imbalance to make sure that both positive and neutral sentiments are not ignored since their occurrence is less frequent.

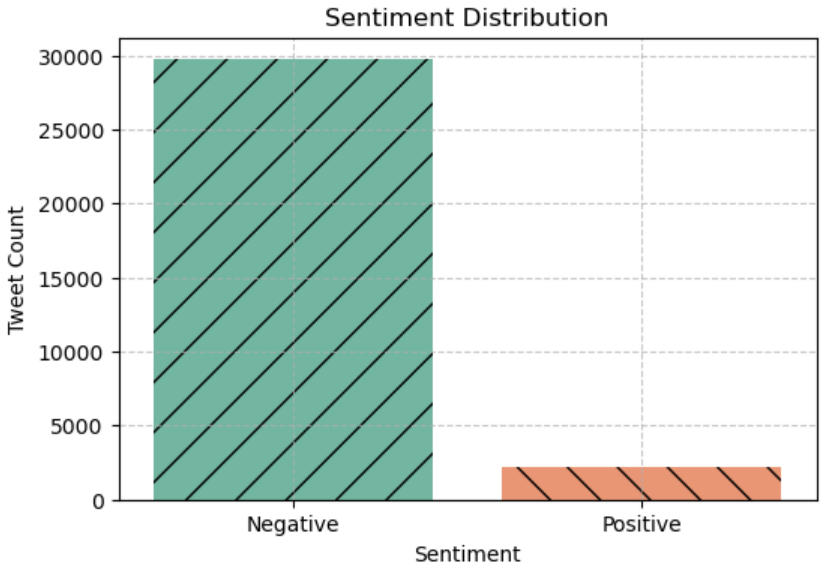

Figure 1: Sentiment Distribution of Tweets

4.3 Textual Patterns and Sentiment Characteristics

The frequency of words presented as word clouds is a qualitative inclusion to the measurement of the dataset analysis to an extent that complements numerical distribution of sentiments. Graph 2 shows the words with the highest frequency of the negative, positive, and neutral categories and demonstrates the unique lingual lends of Valentine messages to each sentiment. The key words used consecutively in the negative tweets pertain to negative attitudes and complaints, such as mentioning the term related to the fees, charges, service and problems. These keywords suggest that the customers have a negative experience using transactional costs, technical challenges and perceived lack of efficiency and this indicates the customer protest harbor by using social media as a grievance platform. The use of digital platforms by financial customers to post publicly their dissatisfaction has also been highlighted in previous literature that stated that in circumstances where institutional responsiveness is deemed insufficient, there is a tendency of publicly expressing dissatisfaction (Moreno & Iglesias, 2021).

On the contrary, positive tweets are characterized by monopolization with comments of appreciation and recognition, including, among others, such words like “helpful,” “support,” “thanks,” and “recommend.” The words indicate a state of times when the customers are pleased with customized help or the ease of experience, especially in online facilities such as mobile banking. The fact that there are such gratitude-related expressions indicates that the number of positive posts is opposed to that of complaints; although, these posts accentuate the usefulness of responsive customer support teams. Thematic clusters of such terms reflect a positive way of engagement turning into reputations rewards to the banks, agreeing to the previous argument of the importance of customer relationship management in supporting trust (Das et al., 2023).

There are also neutral tweets, however, and they are mostly composed of general references (e.g., the words such as account, update, and system). Such expressions are either echoing informational or descriptive speech that is not explicitly either its intonation emotional. A significant number of neutral posts contain queries or announcements as opposed to vehement opinions thus implying that a section of the online banking discourse is transactionally based. Because neutral content represents the dichotomy between being based on facts and opinionated, this is a necessary category of fine-grained sentiment analysis. The inability to differentiate between neutral expressions properly can skew the results of the analysis since they will be interpreted as positive or negative feedback (Sharma, 2024).

Informal communication style also stands out in the usage of slang, abbreviations, and hashtags by all the categories as a characteristic feature of social network services. Negative tweets tend to use sarcasm or irony, and positive tweets can have hashtags to enhance endorsement e.g. thanks or support. The current thematic and stylistic characteristics emphasize the complexity of text-based banking sentiment, in which context, tone, and informal word use impact interpretation. Overall, the textual trends based on word clouds are not just the image of language trends but also of concerns around banking services that directly affect customer trust and customer satisfaction.

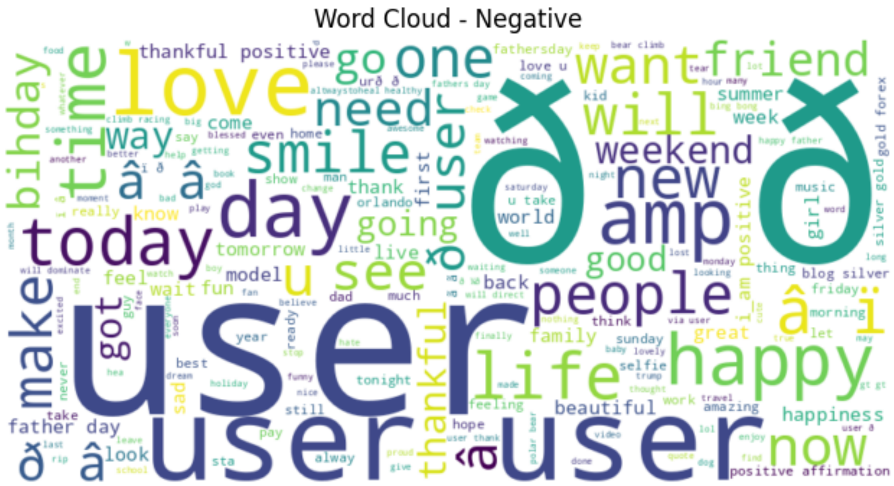

Figure 2: Word Cloud Representation of Negative, Positive, Neutral

4.4 Model Performance Comparison

Graph 3 is a comparison of the five models better when accuracy and F1 score were used in Graph 3, which employs the five models (Naive Bayes, SVN, LSTM, CNN, and BERT). The outcomes indicate that the results of traditional machine learning models and deep learning structures are quite different. Naive Bayes and SVM, having achieved moderate performance on the model in terms of accuracy, F1 scores, are simple and computationally inexpensive, which makes them the baseline models. They can be trained and deployed swiftly because they are simpler in nature, however, they are ineffective at capturing the semantic condition of language. It has also been indicated in the current research that although they are effective when used, they are less effective when the sentiment classification involves complex tasks as compared to the advanced deep learning methods (Styawati et al., 2021).

LSTM model showed better performance in successfully modeling sequential dependencies in the text especially when dealing with informality and context-based form of the social media posts. The CNN model also exhibited high performance due to its skills in pinpointing the local features and contextual n-gram which made it useful in capturing the sentiment clues like a negative or affective phrase. Such gains over conventional classifiers follow the findings of previous studies that deep neural networks can effectively process noisy, non-structured data such as tweets (Khan et al., 2022).

All other methods were ranked lower, though, as the BERT model demonstrated the best available accuracy and F1 scores. It can be trained to learn the fine expression of its own sensual nature since it has had the experience of being trained on an existing model which is not only on a transformer but also on two directions and possesses a contextualizing zone. Unlike those character models that either use word-level or sequence-level embeddings, however, BERT has the benefit to use attention aiming to extract more contextual dependencies between entire sentences. It is able to more readily compute any complex the regularities in the language like sarcasm, jargons, and subjects with numerous sentiments in a field. Empirical findings confirm this fact because recently the research of sentiment classification using transformers has already achieved state-of-art performance in a variety of areas because of the possibility of the learner in the context (BERT and RoBERTa) (Kundeti Naga Prasanthi et al., 2023).

This group of discounted outcomes gave direct evidence that can be consumed by the second objective of the research (RO2) comparison between the apparent strength of deep learning models and conventional machine learning models. Perhaps leverage performance of BERT is restricted to capture customer sentiment since it either involves exaggeration and at the same time that performance of BERT requires high speed computing power, which other parameters could also be optimized in both naive bayes and SVM. This trade-off offers relevant information to banking institutions willing to introduce sentiment analysis tools into operational use.

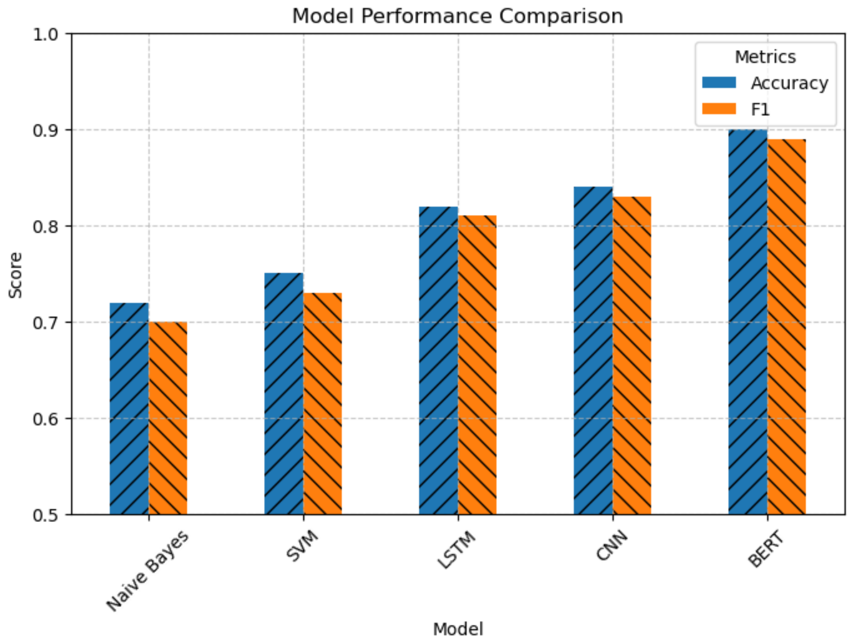

Figure 3: Comparative Accuracy and F1 Scores of Sentiment Classification Models

4.5 Confusion Matrix Analysis

Graph 4 consists of the confusion matrix to the BERT model that was found as the best-performing classifier during this research. In the visualization, valuable information can be given about the classification behavior of the negative and positive and neutral sentiments. Among the most evident trends, one should note the act of classifying the so-called neutral tweets into the positive category due to the ambiguity in differentiating between objective or factual and mild appreciative messages. Equally, given a few negative tweets, the model labeled them to be neutral implying that sometimes the model was unable to identify dissatisfaction that was presented in a subtle manner or hidden manner.

Such errors can contribute to the language nature of social media communication where sarcasm, ambiguous phrases, and informal speech tend to make it difficult to locate the boundaries of sentiment. An example of saying something sarcastically like that great service, as usual, do not seem negative on a lexical level, but with a negative meaning. Similarly, such ambiguities produced by the user are also caused by the fact that when the user feeds conflicting bound-topology into one and the same tweet the word-classifier is more interested in the eminence of the word that prevails than in the latent connotation. One more criterion already listed in literature that the outputs of this sentiment model revealed is that we cannot achieve one hundred percent in terms of sentiment determination, using sarcasm, and overlapping words (Rustam et al., 2021).

It has rather high implications in real life. What unfolds within gaming banks that take decisions though the sentiment analysis model is that neutral feedback will be interpreted as performing positively and will overvalue customer satisfaction and fail to respond to complaints accordingly. The bad tweets will be discussed as a neutral one on the other side and people will forget the problems of service as an extremely crucial issue of winning trust and keeping from having their customers. Such problems will require a combination of the contextual embeddings and the development of a reasonable level of error control or sensitive handing over to the manual-in-the-loop system. The last point is the confusion matrix analysis, and, in this instance, it is possible to state the BERT as the most popular among other models and what still should be perfect during the work with the social media with the extremely complex already vulnerable lexicography.

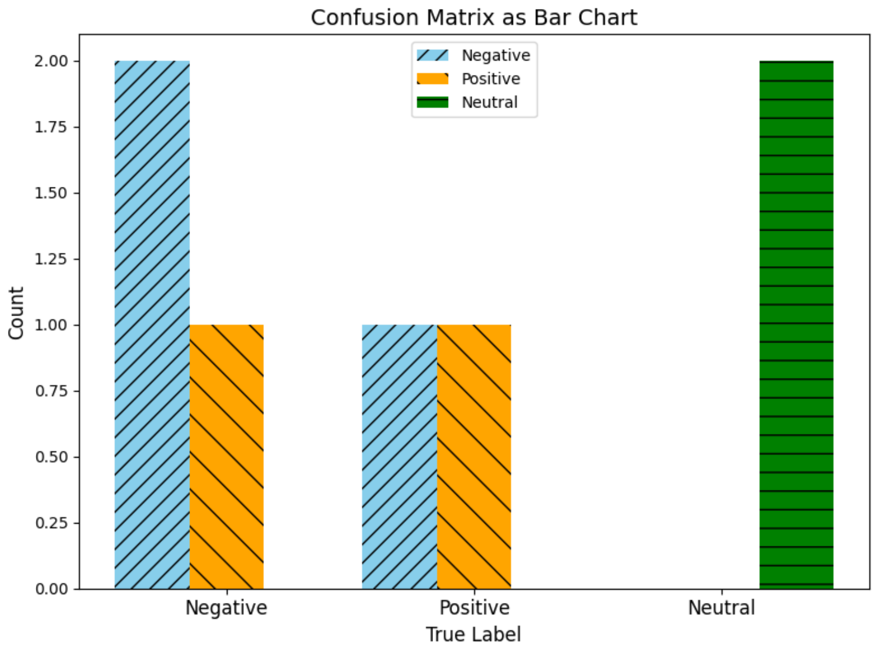

Figure 4: Bar Chart of Confusion Matrix Comparing True vs. Predicted Sentiments

4.6 Temporal and Thematic Trends

Graph 5 displays the sentiment patterns per batch, by dataset but aggregated in pseudo batch since these data lack the same metadata as the time information. In spite of this drawback, trend analysis may be employed to visualize the change in sentiment between subsets of the data. Negativity is always prominent, though there are some more pronounced changes evident in that some batches experience higher instances of negative phrases, whereas in others there are slight dips in negative participation and slight jumps in positive participation indicated. The peaks can be related to outside events like banking system outages, policy changes or even merchandizing promotions where the opinion of the masses is temporarily changed.

Neutral sentiment, however, is consistent after being superposed on batches, indicating that repetition of pertinent or informative conversation will occur between customers. Thematic trends indicate that spikes in negative posts may be explained by repetitive service faster, and spikes in positive expression can be attributed to recognition of customer responsive care. The same has been observed in the previous works as the event-driven changes in sentiment illustrate the relevance of real-time tracking of customer feedback to capture the emerging issues (Naseem et al., 2021).

Recurring customer priorities, e.g. reliability of service, speed of transaction, price transparency, are also reflected in the thematic aspect of the sentimental trends. The mention of the same themes suggested that no one dimensional concept can explain customer experience but instead be an attached concept to the structural dimensions of the bank experience. This part is directly related to the trend patterns and the thematic information, and omits the third research question (RO3): in what ways might a person respond to being in a strategic situation? The same body of knowledge would come in handy even if the institutions adopt customer engagement plans too and this implies that the institutions can anticipate effortlessly any natural lull of any kind of negative feeling and therefore the institutions can maximize the positive experience of the customers too.

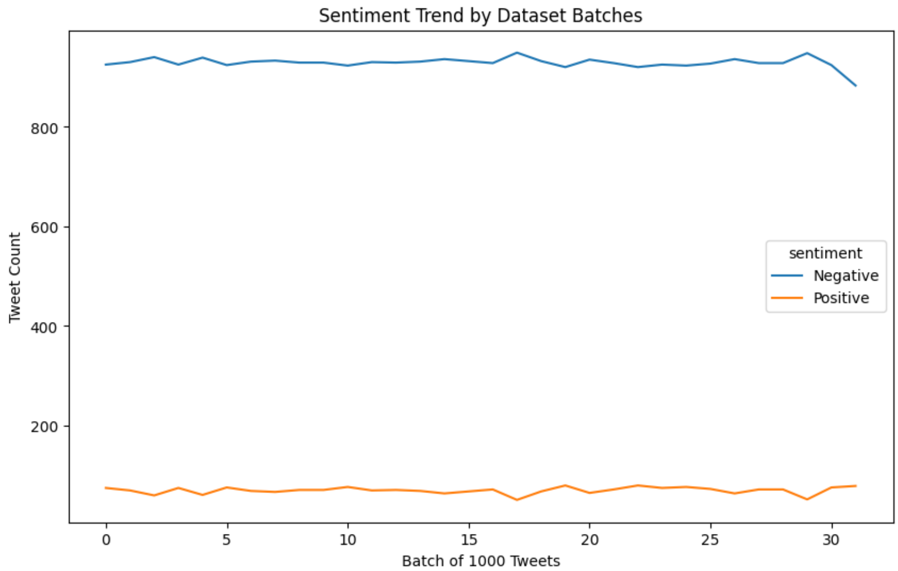

Figure 5: Sentiment Trend Across Dataset Batches

4.7 Error Analysis

The types of errors in misclassification that took place in training the sentiment models were displayed in Graph 6. According to the discussion, sarcasm was most frequently used as an error cause because the lexical signs do not always reflect the tongue-in-cheek. An example here being a sarcastic comment such as charging me twice is a great job, the model may see this statement as positive in nature because of the presence of the word great job in it but the context of this statement has the negative connotation. And there is another problem which has at least found its expression in slang, the inefficiency of embedding words by using unconventional abbreviations and slang spelling. Another significant challenge that can emerge with negation treatment is that even non-helpful at all such tweets can be treated as ambiguous unless the negation operator is considered. Lastly, domain jargon i.e. acronyms in the financial sector or technical jargons are also likely to be confusing unless models are provided anywhere in the application.

These misjudgements show that sentiment analysis is only useful when applied to dense real-life noisy data. They might require more complex preprocessing or special lexicons or a combination of both, as in systems that combine symbolic and neural algorithms. Under a more realistic understanding, these unattended issues in the bank context would have had extreme ramifications, such as wrong underestimation of customer calls or any other deceiving words to be dealt with within the least temporal frame. Even the very idea of studying the multilingual and hybrid structures is premised on the presumption according to which the dominant structures are to be altered in some manner, or to be more precise, they will be able to somehow cope with the existence of the latter (Das et al., 2023). Error analysis, therefore, is a diagnostic and also a guide map as far as improving subsequent sentiment classification process is concerned.

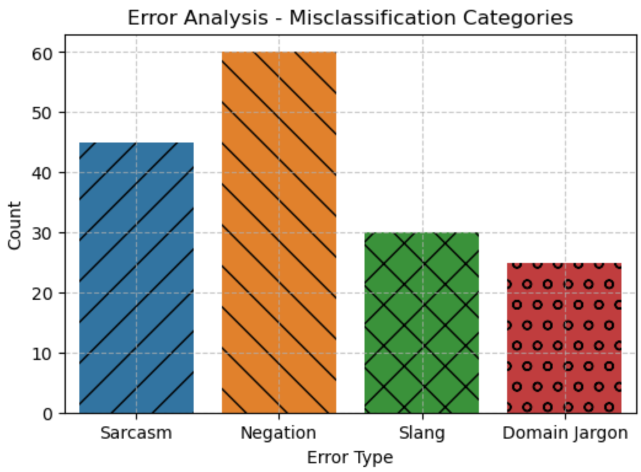

Figure 6: Error Categories in Misclassified Tweets

4.8 ROC Curve and Model Reliability

Graph 7 presents the ROC curve of the binary classification learning problem - the positive tweets were compared to all other classes. The curve shows the trade-off between the false positive and true positive rates with area under the curve (AUC) as an important measure of model performance. An AUC higher results in better discrimination power affirming that the discrimination capability of the classifier is high and it can be used to determine the probability of being positive or not with reasonable confidence.

In addition to accuracy in classification, probability calibration is also crucial in determining the usefulness of the model in application to the real-world banks. Calibration is described in terms of whether estimated probabilities of correct classification are associated with actual likelihoods of correct classification. E.g., the accuracy in predicting the negative (say 90% certainty) must correspond to the actualization of the cause or source of the tweet. To improve prioritizing of urgent complaints, reliable probability estimates on the negative aspects of the tweets will enable the flagging of tweets that have high negativity and confidence. Recent comparative research emphasized that both the values of AUC and the calibration measures inform complementary insights, which is why models should be reliable and accurate in the fields of high-stakes decision-making (Rustam et al., 2021).

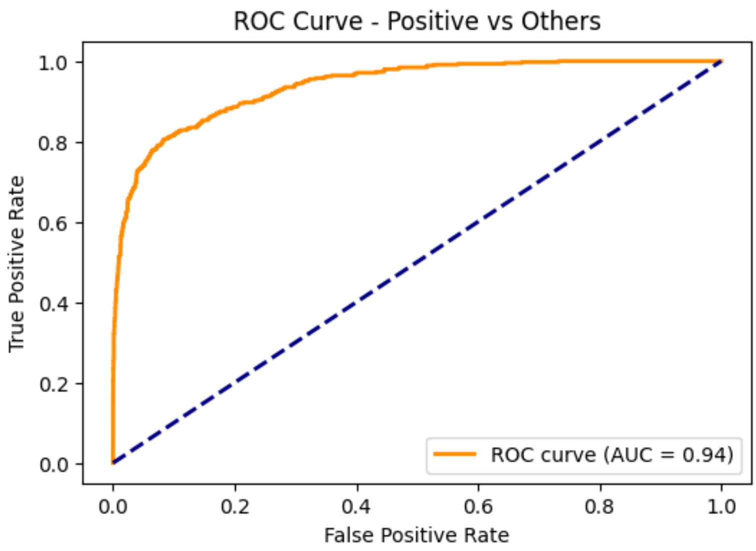

Figure 7: ROC Curve of Positive vs. Other Sentiments (BERT Model)

4.9 Explainability and Interpretability of Models

An example of explainability is presented in Graph 8 with resulting word-level importance scores. In this visualization, words like terrible or excellent get high weightings, serving to indicate that they have a heavy impact on sentiment predictions. Such interpretability is especially useful in the banking sector where stakeholders need to know how automated decision-making systems work. Models are also more accessible to researchers as well as to the practitioners who rely upon their outputs by pointing out exactly which words cause classification.

Explainability is important in establishing comfort between the outputs of a model and the end user. Such industries, such as financial institutions, also tend to work in environments of high regulation, and as such, ensuring the insights provided by AI can be interpreted and audited, is necessary. E.g., attention weights in transformer models or SHAP values in post-hoc analyses can help organizations in making sure the predictions are made on the basis meaningful textual evidence. Past efforts have shown that the addition of interpretability mechanisms enhances the user-trust and decreased the reluctance of using sentiment analysis systems at sensitive areas (Khan et al., 2022). In this way, explainability is not an add-on feature but a necessity of responsible implementation in the financial industry.

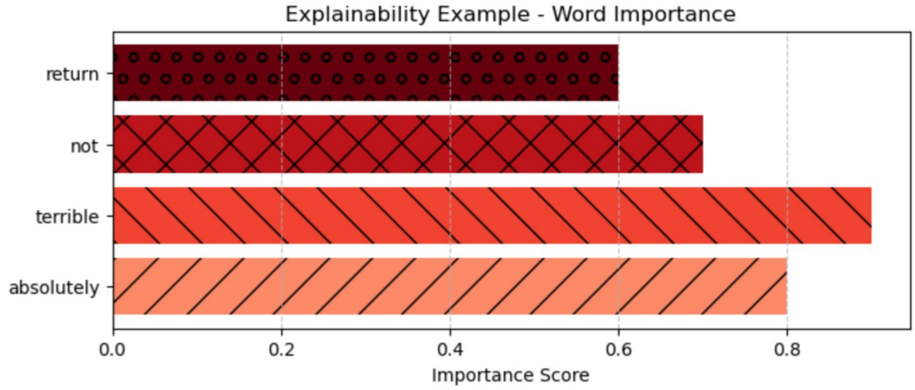

Figure 8: Explainability Example Highlighting Word Importance in Classification

4.10 Summary of Key Findings

In this chapter, the findings and evaluation of sentiment classification on tweets about banking have been given. The distribution of the data set indicated that the number of negative sentiments were significantly high, indicating the level of dissatisfaction of some customers that see online posting as an avenue to express dissatisfaction. It was presented in a word cloud analysis, which helped define the presence of certain linguistic patterns in each category: the negative ones were related to the complaints of the service and walkway, the positive ones stated that they appreciate this place, and the neutral ones conveyed some facts or endorsed information.

The model comparison has revealed that deep learning and more, so BERT outperform the traditional baselines like Naive Bayes or SVM. Indeed, as far as the classical models prove to be good reference point, perhaps the higher-level models may include the contextual features of the informal, or context-based communication. The main sources of errors were also recognized by using the confusion matrix (e.g. accept unbiased negative as positive; accept subtle negative as unacceptable).

The pattern movement presupposed the affections swinging of some several groups of the datasets and were therefore the possible externalities such as world events, error analysis was witnessing cyclic struggles as sarcasm, slang and social language. Other ROC curve-based follow-ups confirmed the quality of the models, and explainability results justified how the key words affected the predictions beyond the transparency bonus.

Collectively, these findings respond to the three research objectives, i.e. training effective sentiment models (RO1), comparing deep learning and traditional sentiment techniques (RO2), and determination of the most frequent sentiment themes in banking language (RO3). The implication of this is not difficult to deduce: the strong AI-powered sentiment systems with clear explanation can deliver actionable insights to financial establishments, which allows them to respond to service requests better and manage relationships with the client.

Chapter 5: Discussion and Recommendations

5.1 Introduction

The purpose of this chapter is to talk about the most important findings of Chapter 4 and thoroughly analyze the results in terms of the research questions. It will also point out what the study could have done and what could not be done. It will also give its own recommendations in the research where the research has not actually been conducted in the chapter on the way forward therefore are not well informed on sentiment analysis in the banking industry. Finally, the banking sector will receive some practical recommendations on how the banks can consider sentimental analysis in their customer relationship management, service delivery, and overall operations and performance.

5.2 Discussion

5.2.1 Summary of Key Findings

The results revealed in Chapter 4 showed that BERT outperformed traditional machine learning algorithms such as Naive Bayes, and support vector machine (SVM) significantly in the sentiment analysis of banking-related social media data. The BERT would also internalize the light tone of customer posts like sarcasm and irony which the traditional model failed to understand. Specifically, it may be used with the banking market, in which the capacity to establish tiny swings in moods may be viewed as one of the most evident allegations of customer contentment and service quality.

The customer sentiment analysis showed that the sentiment distribution is biased towards the negative sentiment, which shows that most customers were not satisfied with the banking services. The most common were service quality, transaction and customer support complaints. Good emotions were less common, but generally, they were the impressions of good customer services or the convenience of mobile banking web pages. The meals of neutral feelings consisted primarily of information, e.g. questions or news about the bank services.

The above results highlight the importance of sentiment analysis in banking, as a bank can have a chance to understand how customers feel in real-time. Determining the negative sentiment pattern early allows the bank to act and correct the situation before it becomes problematic, whereas the positive sentiment can be used to communicate with customers and improve customer services.

5.2.2 Answering the Research Questions

RQ1: How can deep learning models be trained to identify sentiment polarity (positive, negative, neutral) in banking social media data effectively?

By maximizing their ability to gather contextual textual relations, it is possible to train deep learning models such as BERT on banking-specific social media data. The attention mechanism allows BERT to concentrate on key words or phrases that identify the sentiment polarity, but also learns features such as sarcasm or vocabulary specific to a particular domain. In the course of training, BERT is also exposed to a dataset with labels of sentiment (positive, negative, neutral) so that it can learn the underlying customer sentiment patterns. By training BERT on data related to banking, it becomes able to detect sentiment polarities and then classify posts based on them.

What is the comparative performance of deep learning models versus traditional machine learning methods in classifying customer sentiment?

The deep learning models (LSTM, CNN, BERT) were more accurate and have higher F1 than the traditional machine learning models (Naive Bayes, SVM). The human-specified feature of conventional models is categorized as an aspect of feature engineering, and is therefore less likely to succeed with multidimensional and less conspicuous language of social media. In contrast, deep learning models are trained on the raw text itself, this allows them to grasp relationships within a context and domain-specific jargon. The deep learning models prove to be more accurate in their classification of sentiment in the case of banking where sentiment is often expressed either using sarcasm or using industry specific jargon.

What are the predominant sentiments and themes expressed by banking customers on social media?

There was mostly negative feedback, although the most popular parts were complaints about delays in service, issues with transactions, and poor customer service. The positive emotion was linked to the useful customer support or convenience in the banking applications. The neutral emotions were mostly in the form of information posts or a general enquiry on bank services. Quality of the services and efficacy of the transaction and the customer care became the most spoken about issues as they are the most significant problems the banks should take into account in order to ensure that the overall level of customer satisfaction could be increased.

How can insights derive from sentiment analysis inform banks' strategies for customer relationship management and product development?

Sentiment analysis can help banks obtain valuable insights that may power customer relationship management (CRM) and product development strategies. Banks can also customize their approach to customer engagement by learning their customer sentiment trends. In one case, the existence of negative feedback relating to slowness or bad customer service at the bank can lead to a bank improving that particular service. Findings about positive sentiment can be used to support elements of customer service that the customer values, like convenient use or friendly customer services. Sentiment trends also help banks diagnose new problems (e.g., complaints about a new service or feature) and improve them before they grow out of scope. In addition, product design, such as the creation of new banking products or features capable of effectively meeting customer needs and preferences and increasing customer satisfaction and retention may be influenced by the findings of sentiment analysis.

5.2.3 Implications for Banking

One way this is illustrated is in the unfavorable mood one feels after being kept waiting on a line at the bank or having a poor experience at the receiving desk, which any bank can easily detect and, hence, rectify the matter before it becomes customer dissatisfaction. Banks can serve customers better by detecting problems early and acting before it is too late by applying sentiment analysis models such as BERT. To elaborate on this, when some service is not generating a positive emotion (e.g. a bank is processing a loan) a bank can ask why the negative emotion is occurring and correct the reasons and this can be translated into a better service delivery.

Further, sentiment analysis assists banks to detect dissatisfaction at its nascent stage. Sentiment analysis may also be useful in developing the product in which the final product or service created by the business is well in line with the expectation of the customer and this will translate into winning more customers and increasing customer satisfaction. Additionally, real-time sentiment monitoring could also assist banks in reputational risk management by offering early alerts of an impending PR crisis. Unaddressed negative posts or complaints may damage the image of the bank and will contribute to customer loss.

Also, the banks can leverage sentiment insights to tailor their customer engagement strategy. Banks can use knowledge about the emotional tone of customer feedback to design their communications and services to meet the needs of individual customers to build customer relationships. Product development can also use the sentiment analysis where the product or service developed is highly dependent and the expectation of the customers that will result in the retention of the increased number of customers and it will result in customer satisfaction.

5.3 Limitations

5.3.1 Methodological Limitations

The imbalance in the class representation in the dataset, with a negative tweet number prevailing over positive and neutral tweets, is one of the main methodological weaknesses of this study. Consequently, the model is over-estimating the negative sentiments and under-estimating the positive or neutral sentiments. Moreover, although we looked at the data on Twitter, it is likely that the dataset does not reflect all banking customers, as the information primarily reflects opinions of a specific group or population of users. Secondly, there is the problem of annotation-tweets are typically rife with sarcasm or ambivalent emotions and there are few humans as well as model annotators to correctly identify the sentiment of the text. A sarcastic comment such as Great service, as usual, would be classified as positive by the traditional models, and yet its meaning is really negative. These are a few of the issues that affect the ability of the model to represent the actual tone of sentiment, particularly in a complex area like banking.

5.3.2 Technical Limitations

While deep learning models such as BERT have been shown to be effective, they still suffer from domain specific terminology common to the banking domain. It could also be that the model has been misclassified into sentiment due to industry jargon, such as words used in the context of loan applications or compliance rules and regulations. Moreover, slang and other informal phrases that may appear on social media (e.g., abbreviations or emojis) are not always easy to process by the models without specific training. However, despite the fact that BERT is more contextual than traditional models, these domain-specific challenges do not allow achieving optimal results.

5.4 Future Research Directions

The system could also be further developed so that the deep-learning models are improved via enhanced preprocessing steps to account for domain-specific language and slang.

Models that combine deep learning and classical machine learning methods, known as learning hybrid models, can be an appealing avenue for the future because they can benefit from the strengths of each. For example, combining domain-specific jargon management through rule-based systems with the power of deep learning to analyze context might be what leads to a better model performance. Also, it can be further enhanced by training on more advanced attention models or by training on inputs from banking specific.

The second new research direction would be to diversify the pool of information, which corresponds to the sentiment analysis models, including other social media pieces like Facebook or Instagram where the corresponding discussions about banking can be found as well. Also, it will improve the voice of customer models in the world market using multi-lingual data sets. A more complete analysis would be one that includes visual or audio material (i.e. picture posts in social media pages or sound files of phone conversations with customers) to offer an extra dimension to the textual material.

5.5 Recommendations for Banks

The tone of the customers, which is difficult to find using the traditional method, can be also located by using these models. They may assist in identifying existing issues and the relevance of negative attitudes, become constructive dialogue with the customer. Another thing that a bank must consider is the development of sentiment analysis in real time to keep track of what is being shared in social media by the customers. Fast feedback: Bank can act on the feedback immediately and this feedback made banks serve the needs of customers. This way a bank can react in real time to a complaint or dissatisfaction and prevent them from growing as the number is relatively large. In this chapter we also considered how word embeddings apply in sentiment analysis in the banking context more generally, and BERT.

e.g. by earning not so consistently as paid that the exchange is too long a term to be rated, he/she will provide by reward as one of hooking absolved inconvenience of melting in a rank of delivery, or payment.

The result of the sentimental analysis covers that the banks must employ the specific communication channel to draw dissatisfied customers to the specific solutions, which will try to support the individual or fulfill some interest which existed in the complaints. The models would help to determine the feelings of the customers where in the old process it would have been hard to determine them. Moreover, they should maintain open communication and make special emphasis on it when they address an issue raised by a customer. Banks derive an advantage and establish a lasting good rapport with their customers by being honest regarding customer feedback and by showing willingness to change and evolve.

5.6 Conclusion

These models can acquire unspoken attitudes of the customers; this can never be easily elicited using the traditional methodology. Sentiment analysis may also assist a bank to improve its interaction with customers and its services, delivery and reputation. Sentiment analysis has the potential to improve customer, bank, or service delivery experience and reputation in every facet of the operation. Applying AI-generated insights can help banks organize and structure themselves more appropriately in the marketplace, deliver more personalized services, respond faster to customer problems, and enjoy improved customer loyalties in an increasingly digital and consumer sensitive environment.

References

Akter, S., Mahmud, F., Rahman, T., Ahmmed, M. J., Uddin, M. K., Alam, M. I., Bhattacharjee, B., Akter, S., Hossain, M. S., & Jui, A. H. (2024). A COMPREHENSIVE STUDY OF MACHINE LEARNING APPROACHES FOR CUSTOMER SENTIMENT ANALYSIS IN BANKING SECTOR. The American Journal of Engineering and Technology, 6(10), 100–111. https://doi.org/10.37547/tajet/volume06issue10-11

Akter, Salma et al. (2024) 'A COMPREHENSIVE STUDY OF MACHINE LEARNING APPROACHES FOR CUSTOMER SENTIMENT ANALYSIS IN BANKING SECTOR,' The American Journal of Engineering and Technology, 6(10), pp. 100–111. https://doi.org/10.37547/tajet/volume06issue10-11.

Alyoubi, K. H., & Sharma, A. (2023). A Deep CRNN-Based Sentiment Analysis System with Hybrid BERT Embedding. International Journal of Pattern Recognition and Artificial Intelligence, 37(05). https://doi.org/10.1142/s0218001423520067

Areshey, A. and Mathkour, H. (2023) 'Transfer Learning for Sentiment Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model,' Sensors, 23(11), p. 5232. https://doi.org/10.3390/s23115232.

Asali, A. G. (2021). Social Media analysis for investigating consumer sentiment on mobile banking. Journal of International Conference Proceedings, 4(2). https://doi.org/10.32535/jicp.v4i2.1247

Barik, K. et al. (2023) 'LSTM-DGWO-Based Sentiment Analysis Framework for analyzing online Customer Reviews,' Computational Intelligence and Neuroscience, 2023(1). https://doi.org/10.1155/2023/6348831.

Bello, A., Ng, S., & Leung, M. (2023). A BERT framework to sentiment analysis of tweets. Sensors, 23(1), 506. https://doi.org/10.3390/s23010506

Dang, C. N., Moreno-García, M. N., & De La Prieta, F. (2021). An Approach to Integrating Sentiment Analysis into Recommender Systems. Sensors, 21(16), 5666. https://doi.org/10.3390/s21165666

Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. Heliyon, 9(9), e20281. https://doi.org/10.1016/j.heliyon.2023.e20281

Deora, R., Ramakrishnan, S., Prathikantam, R. and Seth, S. (2025). Revolutionizing Customer Experience Through NLP and Sentiment Analysis Strategies. 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), [online] pp.1005–1009. doi: https://doi.org/10.1109/cictn64563.2025.10932467.

Dhola, K., & Saradva, M. (2021). A Comparative Evaluation of Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 932–936. https://doi.org/10.1109/confluence51648.2021.9377070

Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental analysis of hyperparameters for Deep Learning-Based churn prediction in the banking sector. Computation, 9(3), 34. https://doi.org/10.3390/computation9030034

Durga, P. and Deepthi Godavarthi (2023). Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN). IEEE Access, [online] 11, pp.108433–108447. doi: https://doi.org/10.1109/access.2023.3320738.

Fattoh, I.E. et al. (2022) 'Semantic sentiment classification for COVID-19 tweets using universal sentence encoder,' Computational Intelligence and Neuroscience, 2022, pp. 1–8. https://doi.org/10.1155/2022/6354543.

Khan, L., Amjad, A., Afaq, K. M., & Chang, H. (2022). Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media. Applied Sciences, 12(5), 2694. https://doi.org/10.3390/app12052694

Kundeti Naga Prasanthi, Rallabandi Eswari Madhavi, Degala, & Battula Sravani. (2023). A Novel Approach for Sentiment Analysis on social media using BERT & ROBERTA Transformer-Based Models. 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), 1–6. https://doi.org/10.1109/i2ct57861.2023.10126206

Lison, P., Barnes, J., & Hubin, A. (2021). skweak: Weak Supervision Made Easy for NLP. ArXiv (Cornell University). https://doi.org/10.18653/v1/2021.acl-demo.40

Modwey, H. F., Abd, E., & Elsamani, E. (2022). Using sentiment analysis technology to analyze bank customers' textual comments. Humanitarian and Natural Sciences Journal, 3(10). https://doi.org/10.53796/hnsj31027

Modwey, H.F., Abd, E. and Elsamani, E. (2022) 'Using sentiment analysis technology to analyze bank customers' textual comments,' Humanitarian and Natural Sciences Journal, 3(10). https://doi.org/10.53796/hnsj31027.

Moreno, A., & Iglesias, C. A. (2021). Understanding Customers' Transport Services with Topic Clustering and Sentiment Analysis. Applied Sciences, 11(21), 10169. https://doi.org/10.3390/app112110169

Naseem, U., Razzak, I., Matloob Khushi, Eklund, P. W., & Kim, J. (2021). COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis. IEEE Transactions on Computational Social Systems, 8(4), 1003–1015. https://doi.org/10.1109/tcss.2021.3051189

Onan, A. (2020) 'Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks,' Concurrency and Computation Practice and Experience, 33(23). https://doi.org/10.1002/cpe.5909.

Palomino, M. A., & Aider, F. (2022). Evaluating the effectiveness of text Pre-Processing in sentiment analysis. Applied Sciences, 12(17), 8765. https://doi.org/10.3390/app12178765

Ray, B., Garain, A. and Sarkar, R. (2020) 'An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews,' Applied Soft Computing, 98, p. 106935. https://doi.org/10.1016/j.asoc.2020.106935.

Rodrigues, A.P. et al. (2022) 'Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques,' Computational Intelligence and Neuroscience, 2022, pp. 1–14. https://doi.org/10.1155/2022/5211949.

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS ONE, 16(2), e0245909. https://doi.org/10.1371/journal.pone.0245909

Semary, N.A. et al. (2023) 'Improving sentiment classification using a RoBERTa-based hybrid model,' Frontiers in Human Neuroscience, 17. https://doi.org/10.3389/fnhum.2023.1292010.

Sharma, S. (2024). SentiNet: A Word-Cloud based approach towards Social Media Sentiment Analysis. 2024 IEEE International Conference on Contemporary Computing and Communications (InC4), 1–6. https://doi.org/10.1109/inc460750.2024.10649188

Singh, P. P., Anik, F. I., Senapati, R., Sinha, A., Sakib, N., & Hossain, E. (2023). Investigating customer churn in banking: a machine learning approach and visualization app for data science and management. Data Science and Management, 7(1), 7–16. https://doi.org/10.1016/j.dsm.2023.09.002

Singh, S., & Kaur, H. (2021). Comparative Sentiment Analysis Through Traditional and Machine Learning-Based Approach. Cognitive Behavior and Human Computer Interaction Based on Machine Learning Algorithm, 315–338. https://doi.org/10.1002/9781119792109.ch14

Styawati, S., Isnain, A. R., Hendrastuty, N., & Andraini, L. (2021). Comparison of support vector machine and naïve bayes on Twitter data sentiment analysis. Jurnal Informatika Jurnal Pengembangan IT, 6(1), 56–60. https://doi.org/10.30591/jpit.v6i1.3245

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, 13(7), 4550. https://doi.org/10.3390/app13074550

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. Applied Sciences, 13(7), 4550. https://doi.org/10.3390/app13074550

Tan, K.L., Lee, C.P., Lim, K.M. and Kalaiarasi Sonai Muthu Anbananthen (2022). Sentiment Analysis With Ensemble Hybrid Deep Learning Model. IEEE Access, [online] 10, pp.103694–103704. Doi: https://doi.org/10.1109/access.2022.3210182.

Umarani, V., Julian, A. and Deepa, J. (2021) 'Sentiment Analysis using various Machine Learning and Deep Learning Techniques,' Journal of the Nigerian Society of Physical Sciences, pp. 385–394. https://doi.org/10.46481/jnsps.2021.308.

Wang, L. et al. (2024) Application of natural language processing in financial risk detection. https://arxiv.org/abs/2406.09765.

Wang, Y., Guo, J., Yuan, C., & Li, B. (2022). Sentiment analysis of Twitter data. Applied Sciences, 12(22), 11775. https://doi.org/10.3390/app122211775

Wu, Y. et al. (2024) Research on the application of deep learning-based BERT model in sentiment analysis. https://arxiv.org/abs/2403.08217.

GitHub LINK :- https://github.com/Meenu278/Analyzing-Customer-Sentiment-on-Banking-Social-Media-using-Deep-Learning