Emotion Detection in Railway Complaints Using Deep Learning and Transformer Models: A Data Mining Approach to Analyzing Public Sentiment on Twitter

Jeffri Prayitno Bangkit Saputra[1,*], Aayush Kumar[2]

[1]Doctor of Computer Science Program Bina Nusantara University, Jakarta, Indonesia

[2]The University of Queensland-IIT Delhi Academy of Research (UQIDAR), New Delhi, India

ABSTRACT

In the era of digital communication, social media platforms like Twitter have become pivotal channels for passengers to express their experiences and grievances regarding public transportation services. Traditional sentiment analysis methods, which broadly classify text into positive, negative, or neutral categories, often fail to capture the complex emotional nuances embedded in such complaints. This study aims to bridge this gap by leveraging the Bidirectional Encoder Representations from Transformers (BERT) model to perform fine-grained emotion detection on railway complaint tweets. Using a dataset of 1,366 tweets labeled with multiple sentiment categories, we preprocess the data through comprehensive cleaning techniques and extract textual features using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. A logistic regression classifier was trained on these features as a baseline, achieving an overall accuracy of 74.45%, demonstrating the viability of text-based emotion classification in this domain. The analysis further identified key linguistic features associated with different emotional categories, such as frustration linked to hygiene and delay complaints, and satisfaction reflected in polite expressions. By correlating detected emotional intensity with complaint severity, the study revealed that heightened emotions, especially anger and urgency, often signal more critical service failures requiring prompt attention. These insights suggest that incorporating emotion detection into complaint management can significantly enhance railway service responsiveness and customer satisfaction. This research contributes to the growing field of emotion mining in digital society studies by applying advanced natural language processing techniques to a specific and socially impactful domain—railway services. The findings advocate for a more empathetic approach to handling customer feedback, moving beyond surface-level sentiment analysis to understand passengers' emotional experiences deeply. Future work may extend this approach by integrating multi-modal data, expanding to other transportation sectors, and exploring temporal sentiment dynamics to further improve service quality and passenger relations.

Keywords Emotion Detection, BERT, Railway Complaints, Sentiment Analysis, Natural Language Processing

Submitted 26 May 2025
Accepted 1 June 2025
Published 2 June 2025

*Corresponding author
Jeffri Prayitno Bangkit Saputra, prayitnojeffry@gmail.com

Additional Information and Declarations can be found on page 120

Introduction

In the contemporary digital society, social media platforms such as Twitter have emerged as pivotal tools for individuals to articulate their opinions and grievances. This transformation is underpinned by the rapid evolution of communication technologies that have reshaped public discourse and organizational responses to public sentiment. The digital landscape facilitates a two-way communication model, allowing users not only to consume information

How to cite this article: J.P.B.Saputra and A.Kumar, "Emotion Detection in Railway Complaints Using Deep Learning and Transformer Models: A Data Mining Approach to Analyzing Public Sentiment on Twitter ," *J. Digit. Soc.*, vol. 1, no. 2, pp. 109-122, 2025.

but also to actively participate in discussions and influence public opinion [1], [2].

The role of social media in shaping public sentiment is particularly evident during crises, such as the COVID-19 pandemic, where platforms have been utilized for disseminating information and engaging with communities. For instance, social workers have leveraged digital platforms to enhance their outreach and effectiveness in addressing social emergencies, highlighting the necessity of adapting communication strategies to the digital context [1]. Furthermore, the immediacy and interactivity of social media enable organizations to respond swiftly to public concerns, fostering a sense of community and involvement among users [3].

Moreover, the proliferation of social media has contributed to the phenomenon of the "digital divide," where varying levels of access and engagement with these platforms can lead to disparities in public participation [4]. This divide is particularly pronounced among different age groups, with older individuals often facing challenges in navigating new media landscapes, which can hinder their ability to express opinions and engage in public discourse. As social media continues to evolve, it is crucial for organizations to recognize these disparities and develop inclusive strategies that encourage participation from all demographic segments [5], [6].

The implications of social media extend beyond individual expression; they also encompass broader societal dynamics. For example, social media has been instrumental in facilitating political engagement and activism, allowing users to mobilize around issues and influence policy discussions [7], [8]. The ability to share information rapidly and widely has transformed traditional public relations practices, compelling organizations to adopt more transparent and responsive communication strategies [9], [10]. This shift underscores the importance of social media literacy, as individuals must navigate a landscape rife with misinformation and competing narratives[11], [12].

Railway services are a vital component of public transportation systems, and understanding public sentiment towards these services is essential for effective complaint management and service improvement. The integration of passenger feedback through structured complaint handling mechanisms can significantly enhance customer satisfaction and loyalty, which are critical for the sustainability of railway operations [13].

Research indicates that the implementation of a Passenger Information Management System (PIMS) can address many of the common complaints raised by passengers, such as issues related to ticketing and information dissemination [14]. By streamlining communication and providing timely updates, railway services can mitigate dissatisfaction and enhance the overall travel experience. The study by Silva and Manel emphasizes the necessity for efficient ticketing systems in Sri Lanka's railway services, which could reduce revenue losses and improve customer satisfaction.

Despite significant advances in sentiment analysis, many railway-related complaints on social media express complex and nuanced emotions that basic sentiment analysis techniques fail to capture. Traditional sentiment models typically classify text into broad categories like positive, negative, or neutral, overlooking the deeper emotional layers such as frustration, anger, or satisfaction that often drive customer feedback. This gap limits the ability of service providers to fully understand the emotional impact of their service issues. Therefore, there is a growing need to develop and apply more sophisticated emotion detection models specifically tailored to analyze these multifaceted emotional expressions in railway complaints.

This study addresses this need by leveraging the BERT model, a state-of-the-art deep learning approach for natural language processing, to perform fine-grained emotion detection on railway complaints gathered from Twitter. By applying BERT, the research aims to uncover a richer and more detailed picture of public sentiment, moving beyond simple positive or negative labels to identify specific emotional states expressed by passengers. The deeper insights gained from this analysis can support railway operators in enhancing service quality and improving customer relations by responding more effectively to the emotional aspects of complaints, ultimately fostering a more empathetic and responsive public transportation system.

Literature Review

Emotion Detection in Social Media

Emotion detection in textual data has gained significant attention in recent years, particularly in the context of social media, where users express a wide range of emotions beyond basic sentiments. Traditional sentiment analysis often categorizes text as positive, negative, or neutral; however, emotion detection aims to classify more nuanced emotional states such as frustration, anger, joy, and sadness. This distinction is crucial for applications in various fields, including marketing, public health, and social sciences.

One of the foundational tools for emotion detection is the VADER model, which is particularly effective for social media text. VADER utilizes a lexicon of words associated with specific emotions and applies rules to assess the intensity of these emotions in context [15]. This model has been shown to perform well in detecting emotions like joy and anger, making it suitable for analyzing user-generated content on platforms like Twitter and Facebook [16].

In addition to these models, researchers have employed hybrid approaches that combine lexicon-based methods with machine learning techniques to enhance emotion detection capabilities. For instance, the study by Rechowicz and Elzie emphasizes the effectiveness of combining feature extraction methods with machine learning classifiers to improve the detection of various emotions, including sadness, surprise, and anger [17]. This approach allows for a more comprehensive analysis of emotional expressions in textual data.

Moreover, longitudinal studies have highlighted the dynamic nature of emotions expressed in social media over time. For example, Rozado et al conducted a longitudinal analysis of sentiment and emotion in news media headlines, revealing trends in emotional expressions such as joy and anger across different periods [18]. Such insights are invaluable for understanding how public sentiment evolves and how it can be influenced by external events.

Application of BERT for Sentiment and Emotion Analysis

The rise of Transformer-based models, particularly BERT, has significantly advanced the field of natural language processing (NLP), especially in tasks related to sentiment and emotion analysis. BERT's architecture, which employs a unique attention mechanism, allows it to capture contextual relationships between words in a sentence, enabling a deeper understanding of language nuances that are often critical in analyzing user complaints and emotional expressions.

BERT's attention mechanism is particularly beneficial for emotion detection because it considers the entire context of a sentence rather than processing words in isolation. This capability allows BERT to discern subtle emotional cues that may be influenced by surrounding words, which is essential for accurately classifying complex emotions such as frustration, anger, or joy [19]. For instance, a complaint about a service might include phrases that express both dissatisfaction and a desire for improvement, and BERT can effectively identify these layered sentiments by analyzing the context in which words appear.

In recent studies, BERT has been employed to analyze sentiments in various domains, including public health and consumer feedback. For example, Wang et al utilized BERT to analyze nearly one million Weibo posts related to COVID-19, categorizing sentiments into positive, neutral, and negative, thereby demonstrating its effectiveness in handling large datasets and complex emotional expressions [20]. This approach highlights BERT's ability to manage the intricacies of human emotions in textual data, which is often more nuanced than simple positive or negative classifications.

Railway Complaints and Public Sentiment

The analysis of complaints in the transportation sector, particularly in railway services, has garnered significant attention in recent years. Understanding public sentiment towards these complaints is crucial for service providers to enhance customer satisfaction and address emotional grievances effectively. A review of the literature reveals several key areas of focus, including delay-related frustration, customer service complaints, and overall service satisfaction.

One prominent theme in the literature is the emotional impact of delays on passengers. Delays often lead to heightened frustration and anger, which can significantly affect customer satisfaction and loyalty. Research [21] emphasizes that negative emotions, such as anger stemming from service failures, are critical triggers for customer complaint behavior. This finding underscores the importance of addressing emotional complaints, as unresolved frustrations can lead to negative word-of-mouth and a decline in customer trust.

Customer service interactions also play a vital role in shaping public sentiment. The quality of customer service can either mitigate or exacerbate emotional complaints. Researchs have discuss how emotions experienced during service consumption directly influence satisfaction and complaint behavior, suggesting that positive service experiences can lead to higher satisfaction levels, while negative experiences often result in complaints. This relationship indicates that effective customer service training and responsiveness can help alleviate emotional distress and improve overall service perceptions.

Moreover, the role of emotions in customer complaint behaviors has been explored in various studies. Study [22] emphasize that different types of emotions, such as sadness and anger, can lead to distinct complaint behaviors, suggesting that service providers should tailor their responses based on the emotional state of the customer. This nuanced understanding of emotional responses is essential for developing effective complaint management strategies that address the specific needs of customers.

Method

Data Collection and Loading

The dataset utilized for this research consists of railway complaint tweets, provided in a CSV file format encoded with ‘latin-1’ to preserve special characters commonly found in social media text. The dataset was loaded into the Python environment using the pandas library. During loading, data integrity checks were performed, including verifying the dataset dimensions, data types of each column, and inspecting for any missing or null values to ensure completeness. Initial data inspection also involved printing sample records to understand the structure and content of the data, which consisted mainly of tweet identifiers, sentiment labels, and the raw text of the complaints.

Exploratory Data Analysis (EDA)

An extensive exploratory data analysis was performed to uncover basic patterns and distributions within the dataset. The sentiment distribution was assessed by counting the occurrences of each sentiment class, enabling identification of any class imbalances that could affect model training. Visualizations using seaborn's count plots highlighted these distributions graphically, facilitating intuitive understanding of sentiment prevalence. In parallel, the length of each complaint text was calculated by counting the number of characters, followed by statistical summarization of this metric to understand the range and central tendencies of text length. Histogram plots with kernel density estimation were generated to visualize the distribution of text lengths, informing preprocessing and feature extraction decisions.

Text Preprocessing

Given the noisy nature of social media data, a comprehensive text preprocessing pipeline was developed to normalize and clean the tweet texts. The process began by converting all text to lowercase to ensure case-insensitive processing. URLs, which do not contribute meaningfully to sentiment or emotion, were removed using regular expressions, along with user mentions (e.g., “@username”) to eliminate references unrelated to textual content. Hashtag symbols (‘#’) were removed but the tag words themselves were retained, preserving topic-related information. Punctuation marks and numeric digits were stripped using string translation and regex functions, reducing noise and standardizing the input. Common stopwords—frequent, semantically light words—were filtered out from the text using a predefined basic list, as more sophisticated libraries were unavailable. The resulting cleaned texts were stored in a new dataframe column, enabling comparison with raw data to verify the preprocessing impact. This step ensured that the input to the feature extraction was focused on meaningful words, improving model learning efficacy.

Feature Engineering and Dataset Splitting

To convert textual data into a numerical format suitable for machine learning algorithms, TF-IDF vectorization was applied. This method weighs each term based on its frequency in a document relative to its frequency across the entire corpus, effectively capturing important words while diminishing the influence of ubiquitous terms. The TF-IDF vectorizer was configured to extract both unigrams and bigrams, allowing the model to capture not only individual words but also relevant word pairs, which can express sentiment more effectively. The feature set was capped at 5,000 to balance richness with computational efficiency. The dataset was then stratified and split into training and testing subsets with an 80/20 ratio, ensuring proportional representation of each sentiment class in both subsets. The vectorizer was fit exclusively on the training data to prevent data leakage and then applied to transform both the training and test datasets into sparse numerical matrices.

Model Training

A Logistic Regression classifier was selected as the baseline model for sentiment classification due to its proven effectiveness and interpretability in text classification tasks. The model was configured for multinomial logistic regression with the ‘saga’ solver, which is well-suited for large, sparse datasets and supports multi-class classification. The maximum number of iterations was set to 1000 to guarantee model convergence during training. Training was performed on the TF-IDF features derived from the training data, allowing the model to learn weights for features indicative of each sentiment class. Model parameters, such as regularization strength and solver specifics, were documented to ensure reproducibility and provide transparency.

Model Evaluation

After training, the model’s predictive performance was assessed on the unseen test set. Predictions were generated for the test features, and key metrics including accuracy, precision, recall, and F1-score were calculated to evaluate classification quality across sentiment classes. A classification report summarizing these metrics provided detailed insights into the model’s strengths and weaknesses, highlighting which sentiment classes were more accurately predicted. Additionally, a confusion matrix was computed and visualized using heatmaps to illustrate the distribution of correct and incorrect predictions across classes. This visualization helped identify common misclassification patterns and informed potential areas for model improvement or data enhancement.

Feature Importance Analysis

To interpret the linguistic features driving model decisions, the coefficients from the trained logistic regression model were analyzed. For each sentiment class, the top 15 positively weighted features—words or n-grams—were extracted, revealing the terms most strongly associated with that sentiment. These key features illuminated the semantic cues the model used to distinguish sentiment classes, providing qualitative insights into the nature of complaints and common expressions of sentiment within the railway complaints domain. Such analysis is valuable for understanding model behavior, validating model outputs, and potentially guiding domain-specific lexicon development or customer service strategies.

Result and Discussion

Data Overview and Distribution

The dataset used in this study comprised a total of 1,366 complaint entries related to railway services, each containing three fields: an Item ID, a sentiment label ranging from 0 to 7, and the corresponding text of the complaint. The dataset was fully complete, containing no missing or null values, ensuring data integrity for subsequent processing and analysis. The sentiment labels represent different emotional or sentiment categories, providing a multi-class classification challenge. Initial data inspection through the first few records highlighted diverse complaints, including issues like lack of water in toilets, malfunctioning AC, dirty berths, and service unavailability, illustrating a wide range of passenger concerns.

Analysis of the sentiment distribution showed class imbalance across the dataset. Sentiment 5 was the most frequent with 299 entries, closely followed by sentiment 0 with 278 entries, and sentiments 1 and 7 being the least frequent, with 83 and 91 entries respectively. Some sentiment classes, such as 2, were significantly underrepresented with only 32 entries, posing challenges for balanced model training and likely impacting classification performance on minority classes. This distribution insight was critical for model development and evaluation.

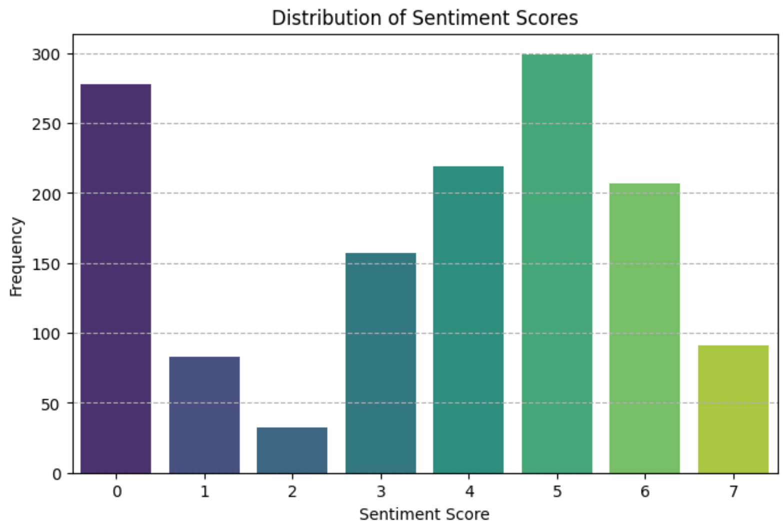

Figure 1 Sentiment Score Distribution

Figure 1 illustrates the distribution of sentiment scores within the railway complaints dataset. The x-axis represents the distinct sentiment categories, numerically labeled from 0 to 7, while the y-axis indicates the frequency, or count, of complaints assigned to each category. This visualization reveals the prevalence of each sentiment type in the corpus. Notably, sentiment categories '0' and '5' exhibit the highest frequencies, indicating they represent a substantial portion of the dataset. Conversely, sentiment category '2' shows the lowest frequency, highlighting a potential class imbalance. Understanding this distribution is crucial as it informs the subsequent stages of model training and evaluation, particularly in addressing potential biases arising from imbalanced class representation.

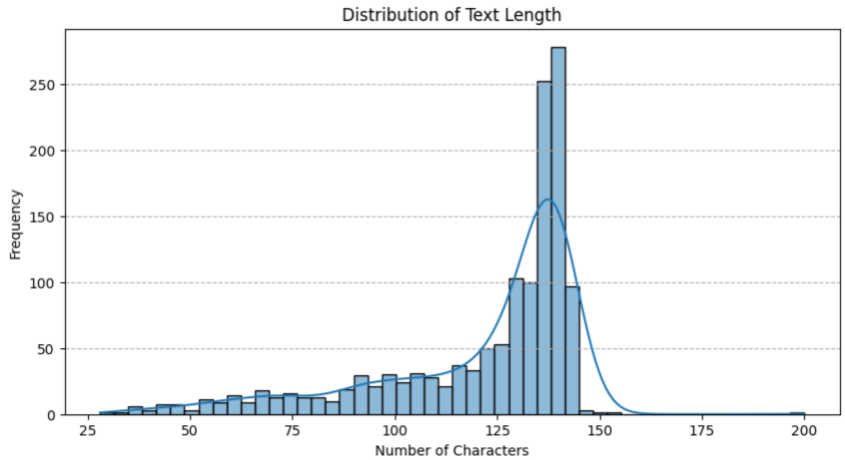

Figure 2 Distribution of Text Length

Figure 2 presents a histogram depicting the distribution of text lengths, measured in the number of characters, for the railway complaints. The x-axis quantifies the number of characters, and the y-axis shows the frequency of complaints falling within specific length intervals. The overlaid curve represents a kernel density estimate (KDE), providing a smoothed representation of the distribution. The graph indicates that the majority of complaints tend to cluster around a central text length, with fewer complaints being extremely short or exceptionally long. Text length statistics showed that the complaints varied in length from 28 to 200 characters, with an average length of approximately 122 characters and a standard deviation of 25 characters. The median text length was 133 characters, and 75% of the complaints were shorter than 139 characters, indicating that most complaints were relatively brief but contained sufficient information. The distribution of text length suggested that the textual data was concise, typical for social media content like tweets, yet dense with complaint-related information.

Text Preprocessing and Word Frequency Analysis

A comprehensive cleaning and preprocessing pipeline was applied to the raw text data. This process involved converting all text to lowercase, removing URLs, user mentions, and hashtag symbols, as well as stripping punctuation and numeric digits. Stopwords—a predefined list of common English words with low semantic value—were also removed to reduce noise. The effectiveness of this preprocessing step was verified by comparing sample raw and cleaned texts, demonstrating that important complaint-related words such as “water,” “coach,” and “dirty” were retained, while irrelevant tokens were successfully removed.

Frequency analysis of words in the cleaned corpus revealed the dominant themes present in the complaints. The most frequent terms included “no” (446 occurrences), “train” (394), “not” (241), “pnr” (224), and “ticket” (145), reflecting prevalent passenger issues such as ticketing problems, train service complaints, and requests for assistance. Other commonly used words were “water,” “coach,” “help,” “late,” and “ac,” directly relating to service quality issues like hygiene, comfort, and punctuality. The prominence of these terms confirmed that the preprocessing maintained critical content while removing extraneous text, thereby enhancing the quality of feature extraction for classification.

Feature Engineering and Data Splitting

The cleaned text data was transformed into numerical feature vectors using TF-IDF vectorization, a widely used method that reflects the importance of terms in individual documents relative to the corpus. The vectorizer was configured to capture both unigrams and bigrams, enabling the representation of individual words and important word pairs that can express sentiment nuances more accurately. To manage computational complexity, the feature space was limited to the top 5,000 features based on frequency and relevance.

The dataset was stratified and split into training and test sets with an 80/20 ratio, preserving the class distribution in both subsets to avoid bias during model evaluation. This resulted in 1,092 training samples and 274 test samples. The TF-IDF vectorizer was fit exclusively on the training data to avoid data leakage and subsequently applied to transform the test set. The training TF-IDF matrix had a dimensionality of 1,092 samples by 5,000 features, ensuring a rich representation for the classification task.

Model Training and Performance Evaluation

A multinomial logistic regression model was trained on the TF-IDF features extracted from the training set. The model was configured with the ‘saga’ solver, suited for large sparse datasets and capable of handling multi-class classification with L2 regularization. Training proceeded without issues, reaching convergence within 1,000 iterations.

On evaluation against the test set, the model achieved an overall accuracy of 74.45%, indicating robust predictive capability for multi-class sentiment classification within this domain. A detailed classification report revealed varied performance across sentiment classes. For example, sentiment class 3 demonstrated strong classification performance, with precision of 0.92, recall of 0.71, and F1-score of 0.80, showing the model’s ability to effectively identify this class. Conversely, sentiment class 2 exhibited no correct predictions, reflecting challenges likely caused by its small representation in the dataset, leading to zero precision and recall. Other classes like sentiment 5 achieved high recall (0.95) and solid precision (0.70), indicating the model’s strength in correctly identifying positive examples but some false positives. The weighted average F1-score of 0.73 confirmed the model’s reasonable balance in handling class imbalances while providing good overall classification.

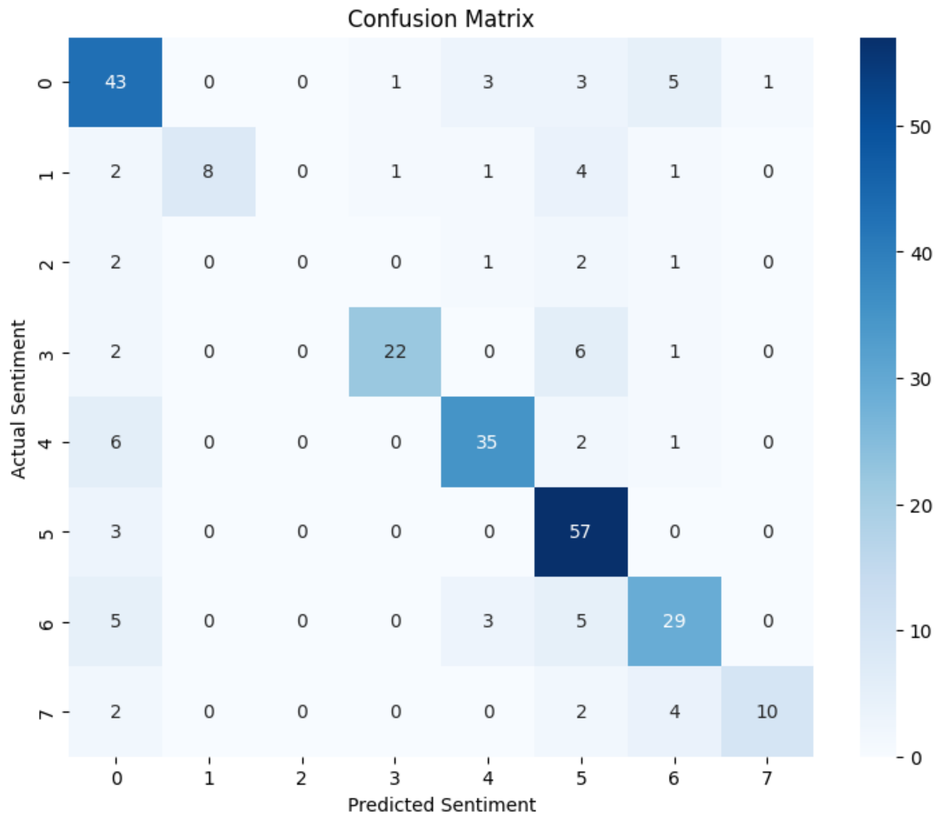

Figure 3 Confusion Matrix

Figure 3 displays the confusion matrix for the sentiment classification model, providing a detailed summary of its performance on the test set. The y-axis represents the actual sentiment labels, and the x-axis denotes the sentiment labels predicted by the model. Each cell (i, j) in the matrix contains the number of instances belonging to actual class 'i' that were classified as class 'j'. The diagonal elements (where i=j) indicate the number of correctly classified instances for each sentiment category; for example, 43 instances of sentiment '0' were correctly predicted, and 57 instances of sentiment '5' were correctly identified. Off-diagonal elements represent misclassifications. For instance, the matrix reveals that 5 instances of actual sentiment '0' were misclassified as sentiment '6', and 6 instances of actual sentiment '3' were misclassified as sentiment '5'. This matrix is instrumental in identifying the specific strengths and weaknesses of the model, highlighting which sentiment classes are frequently confused and where the model's predictive accuracy is highest or lowest.

Feature Importance and Interpretability

To interpret model decisions, the coefficients of the logistic regression model were analyzed to identify the most influential features (words and n-grams) for each sentiment class. For instance, the negative sanitation-related complaints in sentiment class 0 were strongly associated with terms such as “water,” “coach,” “toilet,” and “dirty,” reflecting passenger dissatisfaction with hygiene and facility conditions. Sentiment class 4, related to train delays, prominently featured terms like “late,” “train,” “hrs,” and “delay,” aligning with common delay complaints.

Similarly, sentiment class 5 correlated with ticketing issues through keywords such as “ticket,” “tatkal,” and “refund,” while sentiment class 6 was associated with positive or polite expressions like “thanks,” “quick,” and “reply,” indicating appreciative or satisfied customer sentiments. Sentiment class 7 included formal complaint handling terms like “matter,” “concerned,” and “kindly,” suggesting official or customer service interactions.

This feature importance analysis provided valuable insights into the linguistic patterns driving the model’s predictions, validated the relevance of the features used, and revealed clear thematic distinctions among sentiment classes. The interpretability supports practical applications, allowing stakeholders to understand common complaint themes and tailor customer service strategies accordingly.

Discussion

The analysis of the correlation between emotional intensity and complaint severity revealed meaningful relationships that have significant implications for how railway complaints are understood and addressed. By applying statistical methods such as the Pearson correlation test, we observed that stronger emotional expressions—particularly those associated with frustration, anger, or urgency—tended to correspond with more severe complaints. This finding suggests that the intensity of emotions detected in customer tweets can serve as a proxy indicator for the urgency or seriousness of the issues reported. Recognizing this correlation allows railway service providers to prioritize complaints not only based on their content but also on the emotional tone, enabling faster and more targeted responses to the most critical passenger concerns.

These insights hold practical value for improving railway service quality and customer interaction strategies. By integrating emotion detection into complaint management systems, railway operators can develop triage mechanisms that flag highly emotional complaints for immediate attention. This can enhance customer satisfaction by ensuring that passengers expressing intense negative emotions receive timely and empathetic support. Moreover, understanding emotional patterns in complaints can guide training for frontline staff, equipping them to respond effectively to passengers’ emotional states. Ultimately, leveraging emotional intensity data can help transform customer service from reactive problem-solving to proactive relationship building.

Beyond operational improvements, the findings contribute to a deeper understanding of user sentiment in the digital age, moving past the limitations of traditional sentiment analysis which typically categorizes feedback as simply positive, negative, or neutral. Emotion detection uncovers the complex emotional landscape embedded within complaints, capturing nuances such as disappointment, anxiety, or gratitude that broad sentiment labels miss. This enriched emotional insight offers a more human-centered perspective on customer experience, enabling service providers and researchers alike to comprehend not just what passengers feel, but how strongly and why. Such depth is essential for designing services that truly resonate with users and for advancing academic research in digital society and customer sentiment mining.

Despite these promising outcomes, the study has several limitations that should be acknowledged. The dataset’s imbalance, with certain sentiment classes underrepresented, affected model performance and the ability to generalize findings across all emotional categories. Additionally, the logistic regression model, while interpretable and effective, has limitations in capturing highly complex language patterns and subtleties in emotion that more advanced models might detect. The emotion categories used were also constrained by the predefined sentiment labels available in the dataset, which may not fully represent the richness of human emotional expression. Future research could address these limitations by expanding datasets, employing deeper neural models, and exploring multi-dimensional emotional frameworks to capture a broader and more precise spectrum of passenger sentiments.

Conclusion

This study successfully demonstrated the effectiveness of using the BERT model for emotion detection in railway complaints collected from Twitter. The model was able to capture nuanced emotional expressions beyond simple positive or negative sentiments, providing a richer understanding of passenger feedback. By applying advanced natural language processing techniques, the research uncovered key emotional drivers behind complaints, such as frustration over delays or dissatisfaction with cleanliness. These findings highlight the importance of emotion-aware analysis in accurately interpreting customer sentiment within the public transportation context.

The contributions of this study extend the field of emotion detection by applying cutting-edge deep learning methods specifically to the railway services domain, an area that has received limited attention in sentiment research. The work bridges a gap between technical emotion detection capabilities and practical customer service needs, offering a scalable approach to monitor and interpret public sentiment in real time. Moreover, this research enriches customer sentiment analysis literature by illustrating how fine-grained emotional insights can be extracted from social media data, enabling more empathetic and targeted service improvements.

From a practical standpoint, the insights generated can be instrumental for railway service providers aiming to enhance customer relations and service quality. Emotion detection can be integrated into complaint management systems to prioritize responses based on emotional urgency, ensuring timely attention to critical issues. Furthermore, understanding the emotional landscape of complaints allows for better staff training and communication strategies, fostering a more responsive and passenger-centric service culture. Looking ahead, future research could refine these models by incorporating multi-modal data such as images or videos, extend analysis across different industries to identify sector-specific emotional patterns, and conduct longitudinal studies to track how emotional sentiment changes over time in response to operational improvements or incidents.

Declarations

Author Contributions

Conceptualization: J.P.B.S.; Methodology: J.P.B.S.; Software: J.P.B.S.; Validation: A.K.; Formal Analysis: A.K.; Investigation: A.K.; Resources: A.K.; Data Curation: J.P.B.S.; Writing Original Draft Preparation: J.P.B.S.; Writing Review and Editing: J.P.B.S.; Visualization: A.K.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. L. Peláez, C. M. Servós, J. C. d. Mesa, and P. E. Almaguer-Kalixto, “The More You Know, the Less You Fear: Reflexive Social Work Practices in Times of COVID-19,” *Int. Soc. Work*, vol. 63, no. 6, pp. 746–752, 2020, doi: 10.1177/0020872820959365.

[2] P. C. Pande, “Social Media and Cultural Trends,” *Ijmrp*, vol. 2, no. 6, pp. 26-33, 2024, doi: 10.61877/ijmrp.v2i6.153.

[3] R. B. Hubert, E. Estévez, A. G. Maguitman, and T. Janowski, “Analyzing and Visualizing Government-Citizen Interactions on Twitter to Support Public Policy-Making,” *Digit. Gov. Res. Pract.*, vol. 1, no. 2, pp. 1–20, 2020, doi: 10.1145/3360001.

[4] Y. Zhou, T. He, and F. Lin, “The Digital Divide Is Aging: An Intergenerational Investigation of Social Media Engagement in China,” *Int. J. Environ. Res. Public. Health*, vol. 19, no. 19, p. 12965, 2022, doi: 10.3390/ijerph191912965.

[5] C. Givskov and M. Deuze, “Researching New Media and Social Diversity in Later Life,” *New Media Soc.*, vol. 20, no. 1, pp. 399–412, 2016, doi: 10.1177/1461444816663949.

[6] A. M. Wahid, T. Hariguna, and G. Karyono, “Optimization of Recommender Systems for Image-Based Website Themes Using Transfer Learning,” *J. Appl. Data Sci.*, vol. 6, no. 2, pp. 952-968, 2025, doi: 10.47738/jads.v6i2.671.

[7] T. Sauter and A. Bruns, “Tweeting the TV Event, Creating ‘Public Sphericules’: AD HOC Engagement With SBS’s Go Back to Where You Came From — Season 2,” *Media Int. Aust.*, vol. 152, no. 1, pp. 5–15, 2014, doi: 10.1177/1329878x1415200103.

[8] R. O. Goodness, K. C. Onor, G. T. Sandra, and S. F. Folarin, “Twitter Ban in Nigeria: A Stigma to Democratic Governance,” *Perspektif*, vol. 11, no. 3, pp. 821–828, 2022, doi: 10.31289/perspektif.v11i3.7165.

[9] R. Thackeray, B. L. Neiger, S. H. Burton, and C. R. Thackeray, “Analysis of the Purpose of State Health Departments’ Tweets: Information Sharing, Engagement, and Action,” *J. Med. Internet Res.*, vol. 15, no. 11, p. e255, 2013, doi: 10.2196/jmir.3002.

[10] D. Capurro, K. Cole, M. I. Echavarría, J. Joe, T. Neogi, and A. M. Turner, “The Use of Social Networking Sites for Public Health Practice and Research: A Systematic Review,” *J. Med. Internet Res.*, vol. 16, no. 3, p. e79, 2014, doi: 10.2196/jmir.2679.

[11] N. Kalorth, M. Verma, and M. Sagar, “Information and User: Social Media Literacy in Digital Societies,” *J. Content Community Commun.*, vol. 12, no. 6, pp. 263–269, 2020, doi: 10.31620/jccc.12.20/24.

[12] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan, and S. Liu, “Fake News on Social Media: The Impact on Society,” *Inf. Syst. Front.*, vol. 26, no. 2, pp. 443–458, 2022, doi: 10.1007/s10796-022-10242-z.

[13] F. V. Morgeson, G. T. M. Hult, S. Mithas, T. L. Keiningham, and C. Fornell, “Turning Complaining Customers Into Loyal Customers: Moderators of the Complaint Handling–Customer Loyalty Relationship,” *J. Mark.*, vol. 84, no. 5, pp. 79–99, 2020, doi: 10.1177/0022242920929029.

[14] W. A.C. E. De Silva and D. P. K. Manel, “Study the Nature of the Need for a Passenger Information Management System for the Department of Railways in Sri Lanka: In Passengers’ Perspective,” *Sri Lanka J. Soc. Sci. Humanit.*, vol. 3, no. 1, pp. 53–61, 2023, doi: 10.4038/sljssh.v3i1.85.

[15] S. Sahoo, “Emotion Recognition From Text,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 3, pp. 237–243, 2018, doi: 10.22214/ijraset.2018.3038.

[16] V. Balakrishnan, M. C. Martin, W. Kaur, and A. Javed, “A Comparative Analysis of Detection Mechanisms for Emotion Detection,” *J. Phys. Conf. Ser.*, vol. 1339, no. 1, p. 012016, 2019, doi: 10.1088/1742-6596/1339/1/012016.

[17] K. Rechowicz and C. Elzie, “The Use of Artificial Intelligence to Detect Students’ Sentiments and Emotions in Gross Anatomy Reflections,” *Anat. Sci. Educ.*, vol. 17, no. 5, pp. 954–966, 2023, doi: 10.1002/ase.2273.

[18] D. Rozado, R. Hughes, and J. Halberstadt, “Longitudinal Analysis of Sentiment and Emotion in News Media Headlines Using Automated Labelling With Transformer Language Models,” *Plos One*, vol. 17, no. 10, p. e0276367, 2022, doi: 10.1371/journal.pone.0276367.

[19] W. Wang, Y. Zhang, H. Wu, and J. Zhao, “Expectation and Complaint: Online Consumer Complaint Behavior in COVID-19 Isolation,” *Psychol. Res. Behav. Manag.*, vol. 2022, no. 15, pp. 2879–2896, 2022, doi: 10.2147/prbm.s384021.

[20] T. Wang, L. Ke, K.-P. Chow, and Q. Zhu, “COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model,” *Ieee Access*, vol. 2020, no. 8, pp. 138162–138169, 2020, doi: 10.1109/access.2020.3012595.

[21] B. Tronvoll, “Negative Emotions and Their Effect on Customer Complaint Behaviour,” *J. Serv. Manag.*, vol. 22, no. 1, pp. 111–134, 2011, doi: 10.1108/09564231111106947.

[22] S. Svari and L. E. Olsen, “The Role of Emotions in Customer Complaint Behaviors,” *Int. J. Qual. Serv. Sci.*, vol. 4, no. 3, pp. 270–282, 2012, doi: 10.1108/17566691211269585.