

Service Quality Evaluation in Electronic Invoicing: Sentiment Analysis of Customer Behavior

Ginger M. Machado Ruiz^(a), Angelo M. Silva Pincay^(a), Juan C. García Plúa^(a), Michelle A. Varas Chiquito^(a)

^(a) Facultad de Ingeniería Industrial, Universidad de Guayaquil. Guayaquil, Ecuador, 090114

Corresponding author: gmelanie.machador@gmail.com

Vol. 04, Issue 03 (2025):
July-December
DOI: 10.53591/easi.V4i3.2393
ISSN-e: 2953-6634
ISSN Print: 3073-1526
Submitted: August 08, 2025
Revised: November 13, 2025
Accepted: December 31, 2025

Engineering and Applied
Sciences in Industry
University of Guayaquil. Ecuador
Frequency/Year: 2
Web:
revistas.ug.edu.ec/index.php/easi
Email:
easi-publication.industrial@ug.edu.ec

How to cite this article:

Machado Ruiz, G. M., et al. (2025). Service Quality Evaluation in Electronic Invoicing: Sentiment Analysis of Customer Behavior. *EASI: Engineering and Applied Sciences in Industry*, 4(3), 28-39.
<https://doi.org/10.53591/easi.V3i2.2393>

Articles in journal repositories are freely open in digital form. Authors can reproduce and distribute the work on any non-commercial site and grant the journal the right of first publication with the work simultaneously licensed under a CC BY-NC-ND 4.0.

Abstract. Electronic invoicing has become a fundamental part of the digital transformation of organizations, facilitating operational efficiency. This study evaluates the quality of service in electronic invoicing systems in Ecuador through sentiment analysis applied to user comments. Using advanced natural language processing (NLP) techniques, the BERT model was implemented to automatically classify customer opinions into positive, negative, or neutral sentiments. The analysis was conducted in four stages: data processing, sentiment polarity classification, comment categorization (into recommendations and non-recommendations), and result interpretation. The findings reveal thematic patterns that reflect both user-valued elements and sources of dissatisfaction, particularly in terms of usability, technical support, and system performance. The results show a predominance of positive sentiment, while also highlighting critical areas that need improvement. The model demonstrated high accuracy, enabling real-time visualization of customer perception. As a recommendation, the study proposes optimizing the user interface, automating key processes, and providing staff training to enhance the user experience and strengthen service competitiveness within the digital sector.

Keywords: electronic invoicing; natural language processing; sentiment analysis; service quality.

1. INTRODUCTION

In the Ecuadorian business environment, electronic invoicing has experienced remarkable growth over the past decade, becoming a central pillar of organizations' digital transformation processes. According to Rivas et al. (2024), since its implementation—promoted by the Internal Revenue Service (Servicio de Rentas Internas, SRI)—this system has been widely adopted by companies not only to ensure compliance with tax regulations but also to optimize administrative processes and enhance transparency in their operations. Likewise, electronic invoicing has enabled companies to integrate their accounting systems with online platforms, facilitating more efficient management of tax-related information (Montero, 2023). According to data from the National Institute of Statistics and Census (INEC, 2022), more than 90% of companies in Ecuador have implemented electronic invoicing.

In the digital services sector, particularly in electronic invoicing, service quality is a key factor that directly influences both customer satisfaction and companies' operational effectiveness. This factor becomes especially relevant in a competitive environment, where organizations seek to differentiate themselves through reliable, fast, and user-oriented technological solutions (Franco and Prats, 2021).

Within this context, sentiment analysis is an advanced natural language processing (NLP) technique used to identify, classify, and extract opinions, emotions, and attitudes from textual data. In the field of service evaluation, this tool has become an invaluable resource for companies, enabling them to gain a more accurate and detailed understanding of customers' perceptions, needs, and experiences (Campos et al., 2024).

Furthermore, studies on service quality in this sector have employed various evaluation and improvement models, among which the SERVQUAL model has stood out due to its comprehensive approach to assessing customer perceptions across five essential dimensions: tangibility, reliability, responsiveness, assurance, and empathy (Dueñas et al., 2023).

Similarly, Albán and Gualoto (2024) argue that advanced NLP techniques applied to emotion analysis—driven by large volumes of data and supported by sophisticated architectures—offer higher accuracy and an improved ability to capture linguistic nuances.

The objective of this research is to conduct a sentiment analysis of customer feedback to identify aspects that can be improved, thereby enhancing the quality of electronic invoicing services. This approach aims to strengthen service perception, foster customer loyalty, and reinforce the company's competitive position within Ecuador's electronic invoicing sector.

2. METHODOLOGY

This project was developed through four key phases. The first phase, Data Exploration and Preprocessing, involved preparing the working environment, loading the dataset containing customer comments, and examining its structure. Potential missing values were identified and handled; a qualitative exploration of text samples was conducted, and visualization techniques, such as word clouds and frequency term counts, were applied. Additionally, textual preprocessing procedures—including lowercase conversion, punctuation removal, tokenization, stopword removal, and lemmatization—were used to prepare the corpus for analysis (Pota et al., 2021).

The second phase, Emotional Polarity Classification, employed the BERT model to classify customer comments into sentiment categories: positive, negative, or neutral. The model was trained and fine-tuned over multiple epochs until satisfactory performance was achieved (Vizcaíno & Aguaded, 2020).

The third phase, Comment and Recommendation Classification, focused on refining the categorization of comments based on their usefulness and content, distinguishing between recommendations, non-recommendations, and nonsensical texts (Arango & Osorio, 2021).

Finally, during the Results Analysis phase, the classified comments were interpreted, thematic patterns were identified using topic modeling techniques, and visualizations were generated to represent the findings. This process enabled the extraction of relevant conclusions regarding users' experiences with the electronic invoicing system (Salinas & Díaz, 2021).

2.1. Functional and non-functional requirements

Subsequently, Table 1 presents a breakdown of the functional requirements, which were essential for defining the criteria developed in this study.

Table 1. Functional Requirements

Functional Requirement	Description
Sentiment analysis of comments and reviews	A Natural Language Processing (NLP) model based on the BERT algorithm is integrated to analyze customer opinions and determine their level of satisfaction.
Emotion detection in technical support interactions	NLP techniques are used to identify emotions expressed in customer messages.
Keyword frequency in comments	A TF-IDF (Term Frequency–Inverse Document Frequency) system is employed to identify the most frequently occurring words in customer opinions and to determine the positive and negative aspects of the service.

Source: Machado y Silva (2025).

Accordingly, Table 2 presents the project's non-functional requirements, which were crucial for the management and performance of the selected models.

Table 2. Non-Functional Requirements

Non-Functional Requirement	Description
Periodic updating of NLP models	Sentiment analysis models must be periodically retrained using new customer interactions to enhance accuracy and adapt to shifts in language usage.
Model maintenance and monitoring	MLOps (Machine Learning Operations) techniques are integrated to monitor model performance and automatically adjust models in the event of a decline in accuracy.

Source: Machado y Silva (2025).

2.2. Sentiment Analysis Process

The sentiment analysis process developed in this project aimed to interpret customer opinions regarding the electronic invoicing service using advanced natural language processing techniques. Through several phases, a model was implemented to automatically classify comments based on their emotional polarity, to transform qualitative data into structured information that supports continuous service improvement (Reducindo et al., 2024).

- **Model Selection:** Several models were evaluated, including Naive Bayes, Support Vector Machines (SVM), neural networks (RNN and CNN), and transformer-based models. BERT was ultimately selected due to its superior contextual accuracy.
- **Model Training:** The BERT model was fine-tuned using a previously labeled dataset of customer comments, enabling it to learn patterns associated with different sentiment classes.
- **Model Evaluation:** Performance metrics, including accuracy, recall, and F1-score, were used to validate the model's effectiveness on unseen data.
- **Model Tuning:** Hyperparameters, including the learning rate and the number of training epochs, were adjusted to enhance model performance.
- **Sentiment Prediction:** The trained model was used to classify new customer comments into positive, negative, or neutral sentiment categories.
- **Results Analysis:** Patterns and recurring themes were identified in the classified comments, reflecting customers' perceptions of the electronic invoicing service.
- **Results Visualization:** Graphs, word clouds, and dashboards were generated to facilitate interpretation of the findings.

Finally, Figure 1 presents the general architecture of the sentiment analysis project. It illustrates the main phases, ranging from data exploration to results visualization. This schematic representation helped to organize the workflow in a structured manner and ensured methodological coherence throughout the study.



Figure 1. Overall project architecture. Adapted from the present research.

2.3. Phase 1: Data Exploration and Preprocessing

This phase consisted of preparing the working environment and processing the dataset of customer comments. The required libraries were loaded, the data were imported from Google Drive, and the dataset structure was verified. Subsequently, missing values were analyzed, random comments were reviewed to gain a better understanding of their content, and visualizations, such as word clouds and frequency term counts, were generated. Additionally, text cleaning and transformation tasks were performed, including lowercase conversion, punctuation removal, tokenization, stopwords removal, and lemmatization, to prepare the text for sentiment analysis.

Figure 2 illustrates how the BERT model categorizes customer comments as positive, negative, or neutral. This representation is relevant because it demonstrates the transformation of subjective opinions into emotional categories that facilitate systematic analysis.

```

] #Clasificar automáticamente cada comentario en categorías
#de polaridad básica: positivo, neutral o negativo.

# Cargar pipeline de análisis de sentimiento con modelo BERT preentrenado
sentiment_model = pipeline("sentiment-analysis", model="nlpTown/bert-base-mitmultilingual-un

# Función para mapear etiquetas del modelo a polaridad básica
def map_label(label):
    # Ejemplo: etiquetas del modelo pueden ser '1 star' a '5 stars'
    if label in ['1 star', '2 stars']:
        return 'negativo'
    elif label == '3 stars':
        return 'neutral'
    else:
        return 'positivo'

# Aplicar el modelo a la columna Preguntaa3 y crear nueva columna con etiquetas
df['Asentimiento_Bert'] = df['Preguntaa3'].apply(lambda x: sentiment_model(x)[0][label'])
df['Asentimiento_Bert'] = df['Asentimiento_Bert'].apply(map_label)
print(df[['Preguntaa3', 'Asentimiento_Bert']].head())

```

Figure 2. Text classification. Adapted from the present study.

Figure 3 illustrates the initial distribution of user responses regarding satisfaction and recommendation. The results indicate that the majority of users are willing to recommend the service; however, a group reports dissatisfaction, which serves as an early warning regarding areas that require improvement.

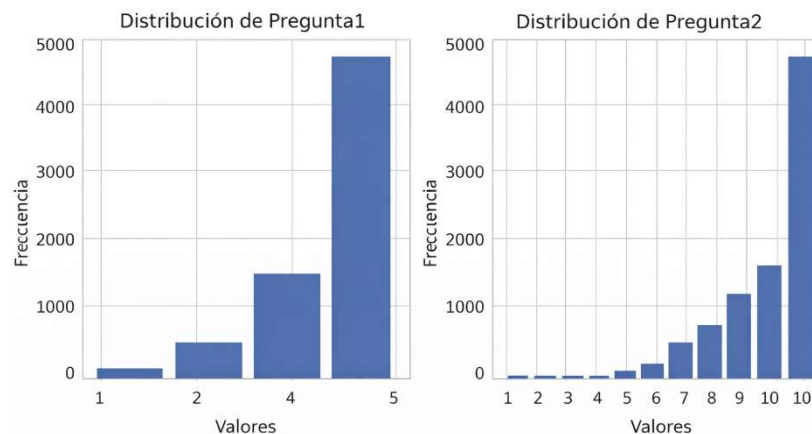


Figure 3. Distribution analysis of Questions 1 and 2. Adapted from the present study.

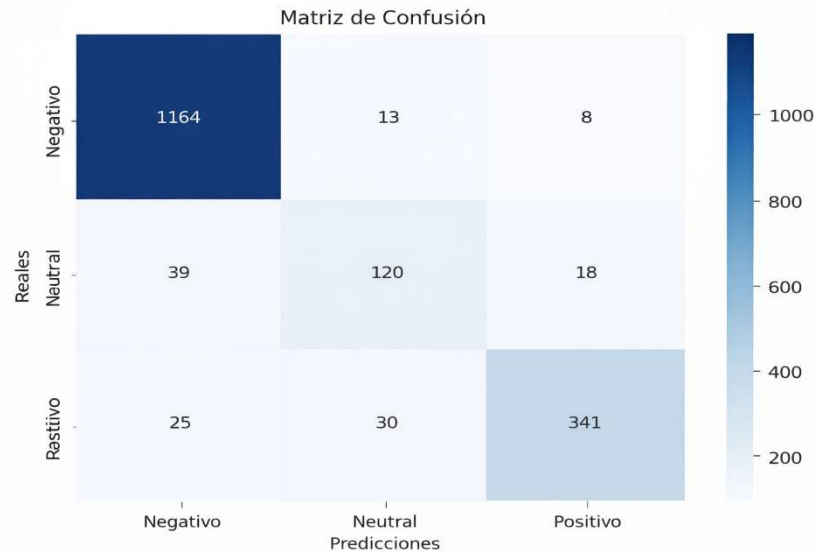


Figure 6. Confusion matrix. Adapted from the present study.

Similarly, Figure 7 illustrates the overall distribution of comments classified as positive, negative, and neutral. A predominance of positive comments is observed, reflecting general acceptance of the service. However, the proportion of negative comments indicates the presence of technical and support-related issues.

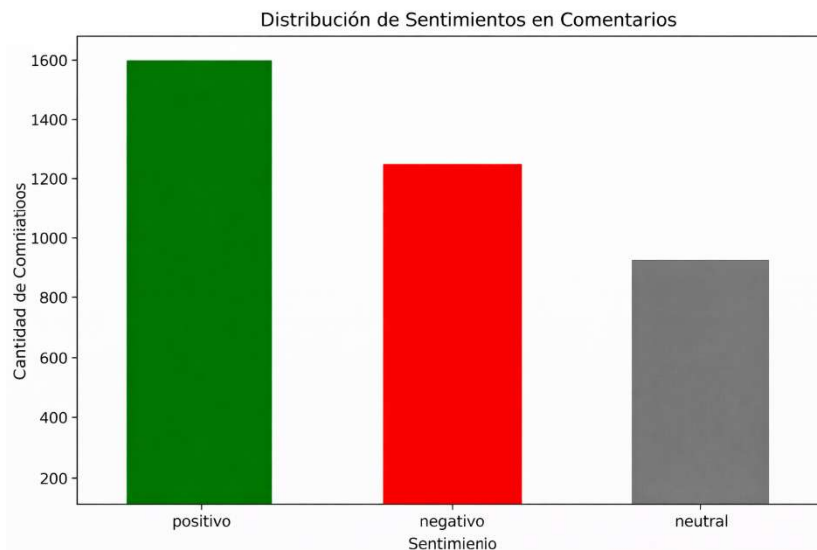


Figure 7. Sentiment distribution of customer comments. Adapted from the present study

2.5. Phase 3: Comment and Recommendation Classification

This phase focused on evaluating user feedback beyond fundamental sentiment analysis by classifying comments into three categories: recommendations, non-recommendations, and nonsensical comments. The BERT model was retrained for this task, achieving solid predictive performance, particularly in accurately identifying negative opinions. Classification errors were analyzed using a new confusion matrix, and the distribution of categories was examined to understand overall trends in user responses.

According to Figure 8, the training results for the model's recommendation classification are presented. Overall performance is satisfactory; however, slight signs of overfitting are observed during validation, indicating a need for parameter adjustments to improve generalization.

```

Epoch 1/3
Training: 100% | ██████████ | 137/137 [00:46:00:00, 2.93it/s]
Evaluating: 100% | ██████████ | 46/46 [00:05:00:00, 8.50it/s]
Train loss: 0.5649 | Val loss: 0.3824 | Val accuracy: 0.8644

Epoch 2/3
Training: 100% | ██████████ | 137/137 [00:50:00:00, 2.72it/s]
Evaluating: 100% | ██████████ | 46/46 [00:05:00:00, 8.76it/s]
Train loss: 0.3207 | Val loss: 0.3469 | Val accuracy: 0.8767

Epoch 3/3
Training: 100% | ██████████ | 137/137 [00:49:00:00, 2.75it/s]
Evaluating: 100% | ██████████ | 46/46 [00:05:00:00, 8.20it/s]
Train loss: 0.2326 | Val loss: 0.4341 | Val accuracy: 0.8425
Test loss: 0.4186 | ██████████ |
Test loss: 0.4486 | Test acc.: 0.8356
    
```

Figure 8. Training results for recommendation classification. Adapted from the present study.

Additionally, Table 3 presents the classification of comments, highlighting high accuracy levels across all classes and a low number of classification errors. These results demonstrate a precise and reliable approach for identifying user intent in customer feedback.

Table 3. Comment Classification Results

Comment Category	Correct Predictions	Classified Comments	Most Common Errors
Non-recommendation	1,232	1,561	33 misclassified as nonsensical
Recommendation	340	1,492	No errors with other classes
Nonsensical	1,243	596	8 misclassified as non-recommendation

Source: Machado y Silva (2025).

The second confusion matrix, shown in Figure 9, presents the classification of comments into recommendation, non-recommendation, and nonsensical categories. The results indicate that the model achieves a high level of accuracy with minimal errors, validating its reliability in interpreting user intent.

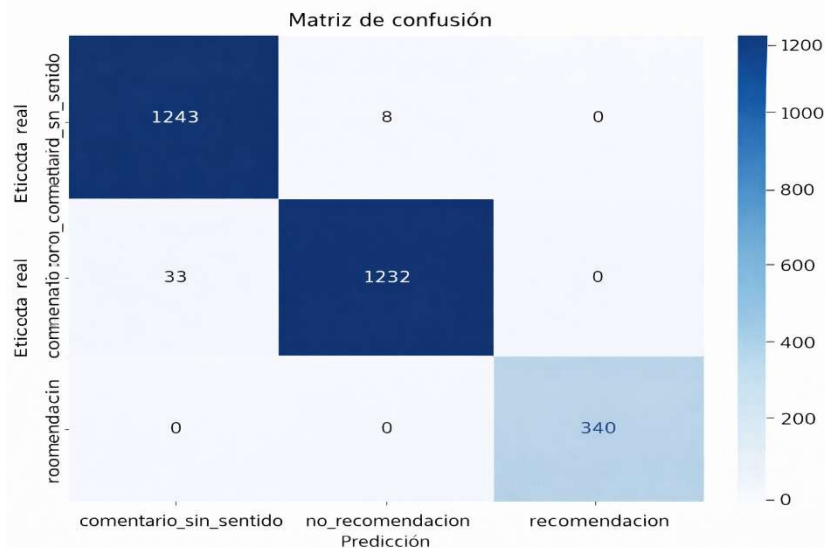


Figure 9. Confusion matrix for comment and recommendation classification. Adapted from the present study.

As illustrated in Figure 10, the graphical distribution of comments classified into each category (recommendation, non-recommendation, and nonsensical) is shown. The results reveal a predominance of negative comments, confirming the need for greater attention to service quality.

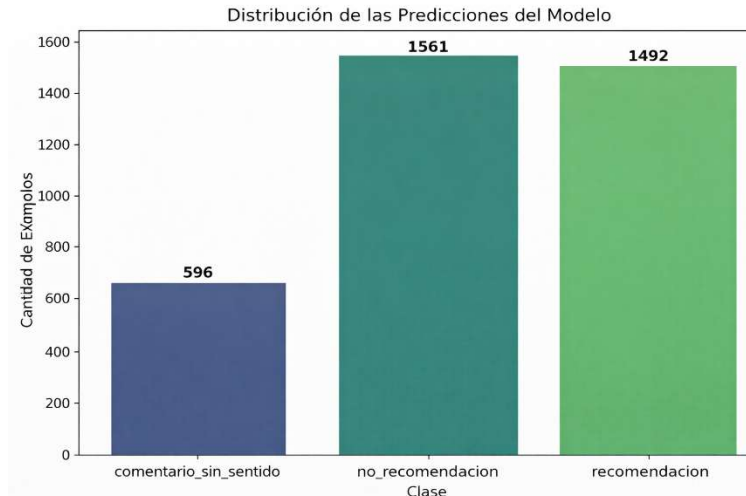


Figure 10. Distribution of model predictions. Adapted from the present study.

3. RESULTS AND DISCUSSION

3.1. Phase 4: Results Analysis

In this final phase, the classified comments were interpreted to gain an in-depth understanding of customer perceptions regarding the electronic invoicing service. The numerical ratings provided by users were analyzed, revealing that the majority expressed satisfaction, although a considerable proportion reflected unfavorable experiences. Through visualizations such as pie charts, word clouds, and frequency graphs, areas with higher acceptance and those generating customer dissatisfaction were clearly identified. In addition, thematic analysis using Latent Dirichlet Allocation (LDA) was applied to uncover recurring topics, allowing the primary sources of dissatisfaction to be categorized into technical issues, usability problems, and unmet expectations.

As shown in Figure 11, the distribution of user ratings, ranging from 1 to 5, is presented. Most ratings are concentrated at higher values, reflecting overall satisfaction; however, a significant percentage falls within lower ranges, indicating unfavorable experiences.

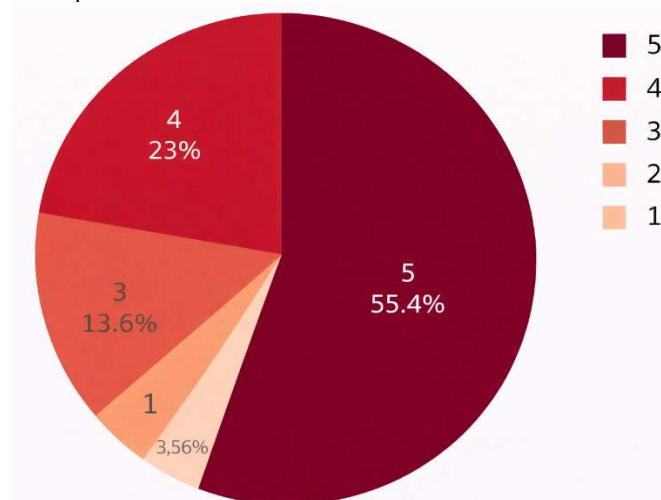


Figure 11. Distribution of customer satisfaction. Adapted from the present study.

Furthermore, the content of positive comments was examined, revealing that many users value system efficiency, process clarity, and ease of use. Positive topic analysis also revealed a balanced distribution across different service areas, indicating that customer satisfaction is not dependent on a single factor, but instead on a combination of well-implemented functionalities. Collectively, these findings provide a comprehensive understanding of the customer

experience and offer key insights for prioritizing strategic improvements that aim to increase satisfaction and customer loyalty.

Figure 12 highlights the most frequent terms in negative comments, such as "error," "slow," and "problem." This finding indicates that the primary sources of dissatisfaction are related to technical failures and response times.

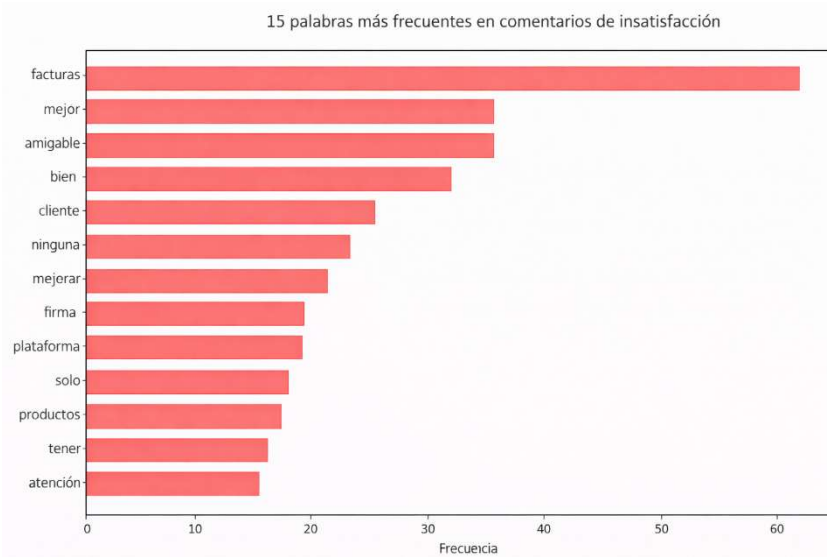


Figure 12. Most frequent terms related to dissatisfaction topics. Adapted from the present study.

Similarly, Figure 13 summarizes the five most common topics identified in negative comments, with a focus on usability issues, technical failures, and unmet expectations. This thematic analysis clearly identifies critical areas requiring prioritized attention.

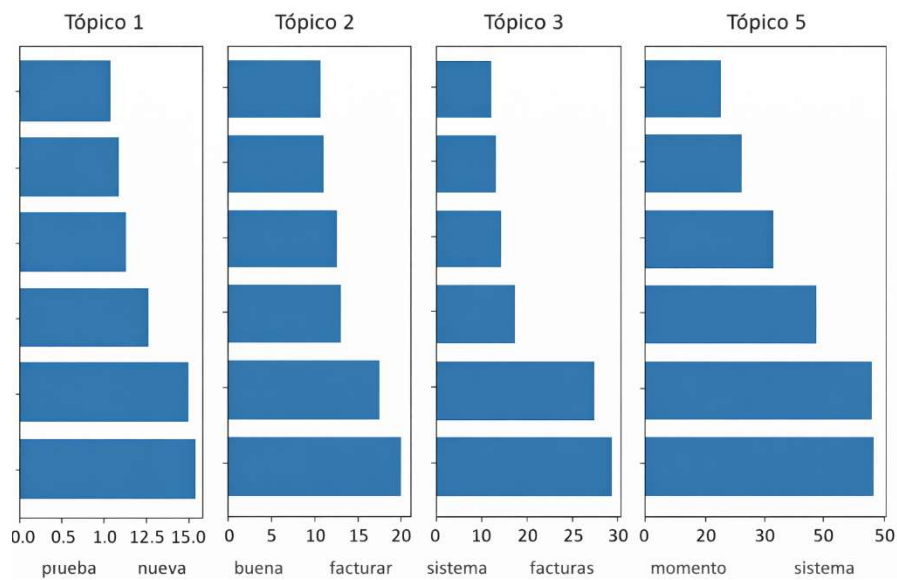


Figure 13. Topic analysis of dissatisfaction-related comments. Adapted from the present study.

Figure 14 provides a global representation of customer satisfaction. The results confirm that, although the service is generally well evaluated, areas for improvement remain, particularly in terms of system stability and technical support.

Distribución de Tópicos en Comentarios Satisfechos

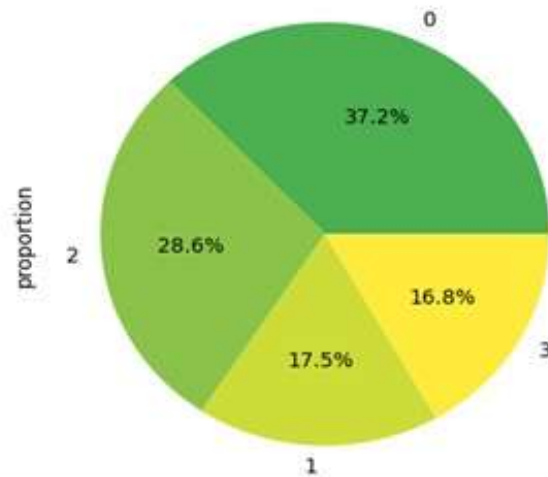


Figure 14. Customer satisfaction overview. Adapted from the present study.

3.2. Discussion

In this section, the findings obtained are compared with those of previous studies, and their practical implications are analyzed. As shown in Figure 7, most of the comments were positive, reflecting a general acceptance of electronic invoicing. This finding is consistent with Franco and Prats (2021), who argue that, within the digital services sector, service quality has a direct impact on customer satisfaction and operational efficiency.

However, a significant volume of negative opinions was also identified, mainly related to technical failures and deficiencies in customer support (see Figure 12). This result aligns with the findings of Rivas et al. (2024), who note that companies have adopted electronic invoicing not only to ensure compliance with tax regulations but also to strengthen transparency in their operations. These results suggest that the perceived quality of electronic invoicing is influenced not only by regulatory compliance and digitalization, but also by system usability and the level of technical assistance provided to users.

Similarly, Figure 13 revealed recurring dissatisfaction topics focused on usability issues and unmet expectations, confirming the need to strengthen customer support mechanisms and improve system stability. On the other hand, positive comments (see Figure 11) highlighted process speed and clarity, which is consistent with Campos et al. (2024), who emphasize that users primarily value simplicity and efficiency in digital services. Likewise, Montero (2023) stresses the importance of speed and efficiency in invoicing processes, as these contribute to cost reduction and environmental sustainability.

Nevertheless, the present study differs from the SERVQUAL-based approach applied by Dueñas et al. (2023), which emphasizes five key dimensions for quality evaluation. In contrast, this study adds value by integrating sentiment analysis supported by advanced NLP techniques, as suggested by Albán and Gualoto (2024). Unlike traditional models, this approach enables the capture of emotional nuances and subjective perceptions, thereby enriching the understanding of customer behavior toward digital services. Moreover, the challenges identified in this study align with difficulties reported in the literature regarding cybersecurity, interoperability, and personalized service, as also noted by Dueñas et al. (2023).

Regarding theoretical implications, the results contribute to strengthening the conceptual foundation concerning the synergy between service quality and customer perception in digital environments. Furthermore, the findings confirm that the SERVQUAL model remains relevant even when sentiment extraction techniques are applied to customer feedback.

From a practical perspective, the findings demonstrate that sentiment analysis can be effectively integrated into the continuous management of electronic invoicing service quality, providing real-time feedback for operational adjustments, user interface improvements, and staff training. This approach addresses the sector's need for immediacy, where trust and user perception are crucial for both customer retention and acquisition, as highlighted by Franco and Prats (2021).

Additionally, another relevant practical implication is that proper management of extracted emotional information enables the anticipation of negative trends—such as technical failures or rising dissatisfaction—allowing for timely interventions to mitigate risk factors, which Campos et al. (2024) identify as a key element in service management.

Among the strengths of this study is the application of advanced NLP and sentiment analysis techniques within a robust methodological framework aligned with best practices. Moreover, the empirical combination of SERVQUAL and deep learning models enhances methodological rigor, allowing for meaningful comparisons with previous literature.

However, it is essential to recognize that the automated interpretation of complex emotions has certain limitations, particularly in texts that are ironic, culturally specific, or ambiguous. Additionally, data representativeness may be biased due to the overrepresentation of extremely positive or highly negative user opinions, which may affect the generalizability of the results. Finally, the lack of a longitudinal evaluation limits the ability to observe changes in service perception over time.

Regarding future research directions, it is recommended to extend the analysis to other sectors and to develop NLP models adapted to Ecuadorian linguistic particularities to achieve more accurate emotional interpretation. Furthermore, conducting longitudinal studies would be highly valuable to examine how customer perception evolves in relation to service quality indicators.

Overall, these findings suggest that although the general perception of the service is favorable, critical gaps remain that must be addressed. The integration of sentiment analysis, as implemented in this study, offers a real-time monitoring mechanism that facilitates early detection of issues, positioning it as a strategic tool for continuous service improvement.

4. CONCLUSION

This study evaluated the quality of electronic invoicing services through sentiment analysis and machine learning models. The use of the BERT model enabled the transformation of textual opinions into valuable data, successfully achieving the project's primary objective.

Additionally, user comments were accurately classified into emotional categories, enabling the identification of both service strengths and areas for improvement. A higher prevalence of positive sentiments was observed; however, criticisms related to technical and operational issues were also detected.

During the analysis, key factors influencing customer experience were identified, including service speed, technical support, ease of use, and the quality of attention received. These elements demonstrated a strong relationship with user satisfaction and loyalty.

Consequently, the BERT model proved effective in capturing customer sentiment in real time, enabling the anticipation of potential dissatisfaction scenarios. This capability opens the possibility of implementing automated monitoring systems to support improved strategic decision-making.

Finally, based on the findings of this study, future improvement initiatives are proposed, including interface optimization, automation of key processes, and staff training, to enhance user experience and strengthen the electronic invoicing service.

Declaration of Conflict of Interest

The authors declare that there is no potential conflict of interest related to the research, authorship, and/or publication of this article.

REFERENCES

- Albán, M., y Gualoto, B. (2024). Desarrollo de un algoritmo para el análisis de sentimientos de textos en kichwa en el ámbito ecuatoriano. Universidad Politécnica Salesiana. <http://dspace.ups.edu.ec/handle/123456789/27234>
- Arango, C., y Osorio, C. (2021). Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning. *Suma de Negocios*, 12(26), 1-13. <https://doi.org/http://doi.org/10.14349/sumneg/2021.V12.N26.A1>
- Campos, J., Ordoñez, E., y Huaylla, A. (2024). Análisis del Sentimiento con NLP en tutoría académica: la vocación como factor de mejora en el rendimiento académico. *C&T Riqchary Revista de investigación en ciencia y tecnología*, 6(2), 7-13. <https://doi.org/10.57166/riqchary.v6.n2.2024.123>

- Dueñas, F., Hidrovo, S., y Loor, I. (2023). Entre el análisis de brechas y el análisis importancia-valoración: una aplicación del modelo SERVQUAL. *Revista San Gregorio*, 1(55), 78-91. <https://doi.org/http://dx.doi.org/10.36097/rsan.v1i55.2388>
- Franco, F., y Prats, G. (2021). Facturación electrónica como herramienta para aumentar la productividad de la empresa. *Investigación & Negocios*, 14(23), 6-16. <https://doi.org/https://doi.org/10.38147/invneg.v14i23.124>
- INEC. (2022). El INEC transparente su gestión de 2022. <https://www.ecuadorencifras.gob.ec/inec-transparenta-gestion-rendicion-cuentas-2022/>
- Montero, D. (agosto de 2023). Módulo de facturación electrónica adaptable al sistema "Atiendo" y evalúo de rendimiento bajo la normativa ISO/IEC 25010 para la empresa PLASTICAUCHO INDUSTRIAL S.A. [Tesis de grado, Universidad Técnica de Ambato]. <https://repositorio.uta.edu.ec/server/api/core/bitstreams/7aef0211-d158-4cb5-a90a-c323fa1de0c9/content>
- Pota, M., Ventura, M., Fujita, H., y Esposito, M. (2021). Multilingual evaluation of preprocessing for Bert-based sentiment analysis of tweets. *Expert Systems with Applications*, 181. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115119>
- Reducindo, J., Calero, H., Fernández, C., y Ramos, E. (2024). Análisis de sentimientos utilizando ChatGPT: una revisión sistemática de la literatura. *Revista Científica Emprendimiento Científico Tecnológico* (5), 1-45. <https://doi.org/10.54798/RWFA3855>
- Rivas, Y., Zheng, K., Romero, T., y Villareal, H. (2024). Gestión administrativa y su impacto en los clientes en el sistema de facturación electrónica de la empresa Ecofiner del cantón Quevedo, año 2023. *Código Científico Revista de Investigación*, 5(1), 1085-1110. <https://doi.org/https://doi.org/10.55813/gaea/ccri/v5/n1/417>
- Sinche Salinas, J. U. Torres Díaz, J. C. (2020). Análisis de sentimientos en los mensajes recibidos en el entorno virtual de aprendizaje de la modalidad abierta y a distancia de la UTPL. Universidad Técnica Particular de Loja. Repositorio Institucional. <https://dspace.utpl.edu.ec/handle/20.500.11962/26503>
- Vizcaíno, A., y Aguaded, I. (2020). Análisis de sentimiento en Instagram: polaridad y subjetividad de cuentas infantiles. *Zer: Revista de estudios de comunicación*, 25(48). <https://doi.org/https://doi.org/10.1387/zer.21454>