Evaluating Word Embeddings for Sentiment Analysis in E-Commerce: A Case Study on MercadoLibre Reviews

Larry Steward Tibaduiza-Mahecha, Ixent Galpin*

Universidad de Bogotá-Jorge Tadeo Lozano, Bogotá, Colombia

Abstract

This study evaluates the effectiveness of various word embedding techniques – Word2Vec, GloVe, MPNet, and ADA-002 – in sentiment classification of product reviews from the e-commerce platform MercadoLibre. Using web scraping, a dataset of over 200,000 user reviews was collected and preprocessed for analysis. The textual data were transformed into vector representations using each embedding method and subsequently fed into four machine learning classifiers: Logistic Regression, Support Vector Machines, Random Forest, and Decision Trees. Performance was measured using accuracy, precision, specificity, and F1-score. Results indicate that ADA-002 embeddings consistently outperform the others across all classifiers, particularly when paired with Support Vector Machines. The findings highlight the value of contextual embeddings for interpreting user sentiment and inform the design of automated opinion mining systems for e-commerce applications.

Keywords

Sentiment Analysis, Word Embeddings, Machine Learning, Natural Language Processing (NLP), E-commerce Reviews

1. Introduction

In an era marked by the exponential growth of digital content, the accurate and timely interpretation of natural language has emerged as a major challenge for computational systems [1, 2]. The inherently complex and unstructured nature of human language poses significant obstacles to automated processing, particularly in tasks such as sentiment analysis, opinion mining, and contextual understanding [3]. In the context of e-commerce, these challenges hinder platforms from effectively capturing user emotions and interpreting product-related feedback, thereby impacting both strategic decision-making and the overall quality of services provided. Indeed, sentiment analysis has become a critical tool for e-commerce retailers: e-commerce sentiment analytics platform Nimble has found that 99% of customers check reviews before purchasing, and 96% specifically seek negative reviews to assess product quality [4].

The application of advanced natural language processing (NLP) techniques, particularly through the use of word embeddings, has shown significant promise in addressing the challenges associated with unstructured textual data. Pre-trained embeddings convert raw text into dense vector representations that capture both semantic and contextual relationships, thereby enabling efficient and scalable analysis of large corpora. Models such as Word2Vec, GloVe, MPNet, and ADA-002 have demonstrated high effectiveness in extracting linguistic meaning, supporting the development of predictive systems capable of accurately evaluating and classifying user opinions [5]. Beyond improving the analytical capabilities of machine learning models, these embeddings also contribute to enhancing user experience on digital platforms by enabling more intelligent and context-aware information processing [6].

This study investigates the application of word embedding techniques and machine learning models in the analysis of product reviews from MercadoLibre[1], the largest e-commerce platform in Latin America [7]. MercadoLibre facilitates online transactions between buyers and sellers, offering a wide range of consumer products accompanied by user-generated feedback. By employing automated data extraction methods, specifically web scraping, this research demonstrates how textual reviews can be

ICAIW 2025: Workshops at the 8th International Conference on Applied Informatics 2025, October 8–11, 2025, Ben Guerir, Morocco

*Corresponding author.

larry.tibaduizam@utadeo.edu.co (L. S. Tibaduiza-Mahecha); ixent@utadeo.edu.co (I. Galpin)

0009-0000-7422-7708 (L. S. Tibaduiza-Mahecha); 0000-0001-7020-6328 (I. Galpin)

[1] https://www.mercadolibre.com.co/

transformed into structured data through embedding-based representations. The resulting vectorized information serves as input for predictive models, thereby enabling the extraction of actionable insights and supporting data-driven decision-making in the e-commerce domain.

The research contributions of this work are:

- To provide a comparative analysis of embeddings, evaluating the effectiveness of Word2Vec, GloVe, MPNet, and ADA-002 embeddings in classifying sentiment in product reviews from MercadoLibre.
- Assessing the performance of multiple classifiers (e.g., SVM, Random Forest) using different embeddings, identifying ADA-002 with SVM as the most effective combination.
- To demonstrate the practical utility of embedding-based sentiment analysis for e-commerce platforms and provide a foundation for future research on embedding dimensionality and model performance.

The structure of this paper is as follows. Section 2 presents the state-of-the-art, with a review of embedding techniques and their applications within natural language processing (NLP), and a critical examination of the ethical and technical challenges inherent in automated textual data extraction. The subsequent sections are broadly based on the widely established CRISP-DM methodology [8], which provides a structured framework for organizing and managing data science projects, guiding the process from business understanding to deployment: Section 3, *Business Understanding*, describes the operational framework of the MercadoLibre platform and details the web scraping methodologies employed to acquire product-related data. Section 4, *Data Understanding*, presents statistical analyses of the collected reviews, emphasizing their significance for the development of predictive models. We subsequently discuss *Data Preparation* in Section 5. Section 6, *Modeling*, presents various text vectorization approaches utilizing different embedding models, alongside their integration with machine learning algorithms. In Section 7, we present a performance comparison using various evaluation metrics, including accuracy, specificity, and F1-score, demonstrating the superior performance of the ADA-002 embeddings when combined with Support Vector Machines. Finally, Section 8 presents conclusions, discussing the practical implications of the developed models for opinion mining tasks and the efficacy of embedding techniques in enhancing NLP predictive capabilities.

2. Related Work

Data science has been transforming natural language analysis with new embedding techniques [9], achieving greater accuracy and performance in text-based models. Xu *et al.* [10] considers embeddings effective in converting high-dimensional data into continuous, dense, and lower-dimensional vector spaces through Gaussian distribution and standard metrics. There is a wide variety of embedding types; however, three dominant techniques are presented: traditional word embedding, static word embedding, and contextualized word embedding [11]. However, Grohe*et al.* [12] argue that they have received little attention from a theoretical perspective, establishing a pre-training approach in machine learning for these techniques.

In state-of-the-art models, embeddings provide greater modeling flexibility, improving parameter allocation during fine-tuning stages [13]. One embedding strategy is word2vec, which encodes words into vectors that facilitate mathematical interpretation by computers for performing natural language classification and regression tasks [14]. Another area for implementing these techniques is social media, where embeddings like GloVe are used to automate the blocking of inappropriate content [15]. Additionally, sentiment analysis is a key goal in natural language analysis, with techniques like MPNet generating vectors containing emotional semantic information [16]. In this way, the semantics and context of words help address language interpretation objectives, automating tasks that are challenging to perform manually.

On the other hand, in a world rapidly advancing through technology, there is a vast array of natural language information to analyze, and web scraping is a tool that automates the extraction of all this data [17]. There are two types of scraping: the first retrieves resources directly from the webpage's code,

and the second uses an interface provided by the page through APIs [18]. Lunn *et al.* [19] emphasizes that web scraping, from a connectivist methodology, is a field of research in computer science and is useful for extrapolating large amounts of data from public sources. Additionally, this web mining technique has an ethical and legal component that generates controversy and cannot be overemphasized when collecting data from websites [20]. Therefore, the development of this mining technique must maintain data integrity and privacy, mitigating the risks of cybercrime [21].

3. Business Understanding

The MercadoLibre platform revolutionized e-commerce, giving rise to new structures of intermediation between markets [6]. Its operation is based on bringing sellers and buyers together through strategies and mechanisms that control, evaluate, and sanction the performance of products.

The purchasing process begins with access to the marketplace portal, where suppliers provide product information, and the consumer can compare based on features, reviews, prices, location, and more. Subsequently, the online purchase is made, representing an economic transaction, which leads to the next step in the chain: payment management. This is done through various payment methods such as debit or credit cards, collection points, or cash on delivery. Once the payment is confirmed, the delivery logistics are initiated, including product shipping, distribution, tracking, and final delivery. The last stage is the post-sale phase, which aims to gather customer feedback, address product-related concerns, and resolve any queries [22].

The feedback provided by the customer after using the product allows other users to trust the product's quality and perceive both positive and negative emotions from the direct experience [23]. Additionally, a numerical rating from 1 to 5 is provided in different aspects such as value for money, durability, among others, which are then summed and averaged to give a score to the product. Subsequently, the MercadoLibre platform takes the products with the highest ratings and presents them in the top search results, a marketing strategy that benefits sellers who justify all the features in their listings.

This sales process on the MercadoLibre platform is accompanied by security measures that protect the B2C relationship. For example, the "Compra Protegida" (Protected Purchase) guarantees the customer 30 days of coverage after receiving the product to assess its quality and voluntarily provide a rating and/or feedback on its functionality, making the feedback transparent and unconditional [24]. Ultimately, this process, with guarantees for all stakeholders, ensures high perception and preference on the platform, generating a large flow of reviews that encompass different types of emotions and technical concepts about the products [25].

4. Data Understanding

The extraction methodology used to obtain data from the MercadoLibre platform is web scraping, a technique used to extract data from websites in an automated manner through the programming of scripts that collect labeled information and store it for subsequent analysis [26]. The information presented on the website is of a commercial nature, identifying the price, average rating, number of reviews, features, seller, and more for each product.

For the dynamics of this scraping, the extraction of available reviews for 7,344 listings in the categories of speakers, televisions, computers, cell phones, keyboards, tablets, headphones, smartwatches, consoles, cameras, and printers is developed. However, only 5,652 products have at least one review. Figure 1 presents a histogram showing the frequency (number of products) across different review ranges. It is observed that over 2,000 products have fewer than 9 customer reviews. The next group consists of those with between 10 and 19 reviews, totaling approximately 800 products, while the third group includes those with more than 200 ratings. In this last category, there are approximately 700 items that were narrowed down, as 3% of them had more than 10,000 reviews.

In a more detailed analysis of the reviews, Table 1 allows us to identify that, in the 1 to 5 rating scale, the highest score accounts for more than 50% of the reviews, while the rating of 2 gathers the

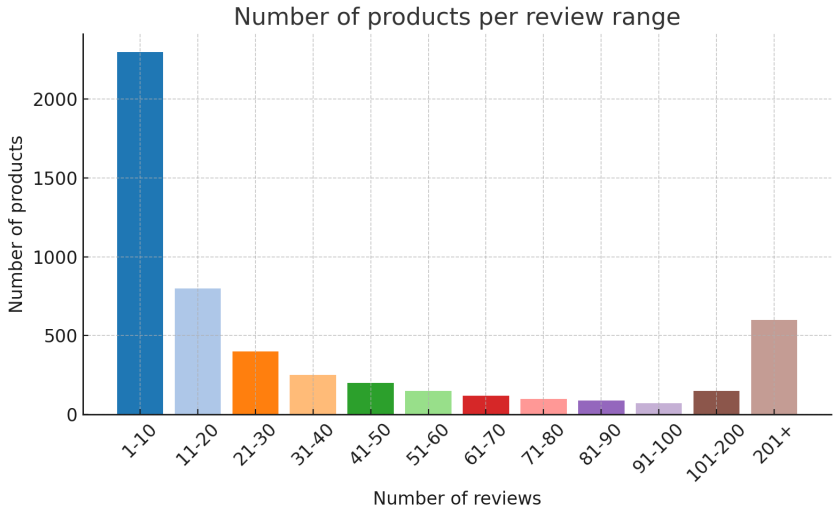

Figure 1: Number of products by review range. To improve clarity, reviews below 100 are shown in bins of 10 (where most variation occurs), while reviews between 101–200 are grouped together. Products with more than 200 reviews are aggregated into a single bin (201+), since they represent a small proportion of the dataset but illustrate the long-tail distribution.

fewest reviews, with 2,233 out of a total of 202,414 comments. This distribution in the ranking reveals that, for the 11 categories studied, the majority of reviews are positive, and the products available on MercadoLibre's platform are well-received by buyers. Sánchez *et al.* [27] determine that in online shopping, consumer trust is directly related to the user conversion rate. Thus, consumer reviews can contribute to the reputation of sellers and provide a competitive advantage.

Table 1
Number of comments by rating

Rating	Number of Comments
1 / 5	4,772
2 / 5	2,233
3 / 5	4,728
4 / 5	23,741
5 / 5	166,940
Total	202,414

Figure 2 presents a more detailed analysis of the stored reviews using a graph visualization, where words that appear at least 80 times and have a minimum weight of 80 connections can be identified. After a process of tokenization, cleaning using stop words, and lemmatization [28], the graph reveals how central words have high frequency, including product themes and some physical features such as "keyboards," "charging," "cables," "headphones," "boxes," among others, accompanied by adjectives like "bad," "excellent," "better," "quality," etc. This highlights that buyers evaluate products both in general and in detail regarding their different components.

The in-degree and out-degree for influential nodes are strong because the length of the edges in most cases is short, leaving on the outer part of the graph the words that, despite having a high frequency, are not closely connected, maintaining a graph without partitions. Additionally, the outer words could help discriminate the different ratings assigned by the consumer, as their occurrences generate segmented groupings that aid in understanding the context of the reviews and reveal information beyond the influential words.

Finally, the data extracted from the MercadoLibre reviews presents a readable and manipulable

Figure 2: Graph visualization of the reviews, filtered by frequency and occurrence

structure for the creation of embeddings, which will then be used as predictive variables in the supervised models to be determined.

5. Data Preparation

Data cleaning begins at the mining stage through web scraping procedures. The information on the MercadoLibre website is distributed both vertically and horizontally. Each page displays approximately 54 products, from which structured information is extracted, including the title of the listing, price, number of reviews, and score (1 to 5 stars), to be stored locally. However, product reviews require a different approach, involving validations through regular expressions to determine which parts of the text box contain the content written by the consumer.

The simplest way to store the information is through JSON structures, which are lightweight data formats and allow for storing large volumes of records within a key-value structure [29]. In this way, all reviews per product are saved along with their ratings, undergoing a transformation process of the

JSON that ensures the review is paired with a rating. If not, that review is excluded.

Following the storage of information, records that are null despite having an associated score are removed from the database. To further refine the characteristics of the reviews, stop words in Spanish are eliminated based on a downloaded dictionary [30]. Each record is converted to lowercase and processed through tokenization using the `nltk` library, resulting in records containing the necessary attributes.

Rating	Frequent Bigrams
1 / 5	poor quality - stopped working - bad product - long battery life - terrible product
2 / 5	poor quality - works well - sound quality - fast discharge - battery life
3 / 5	Long-lasting battery - works well - sound quality - good price - good price
4 / 5	quality price - excellent product - works well - sound quality - battery life
5 / 5	excellent product - excellent product - quality price - excellent quality - works well

Table 2
Frequent bigrams by rating

In the exploratory analysis, the distribution of opinions by rating was identified. It was observed that ratings of 4 and 5 had a high representation, leading to a balancing of records through random sampling. Table 2 shows the five most frequent bigrams for each rating, considering that each rating was balanced to a maximum of 2,000 records. It is evident that ratings below 3 have high frequencies of negative bigrams, while ratings of 4 and 5 include more positive words. Additionally, some bigrams are shared across ratings, such as "short battery life", which indicates that some opinions, despite highlighting negative features, also acknowledge certain positive attributes of the product. This enriches the embeddings by requiring them to capture the context for vectorization. This procedure enhances the models' ability to interpret information and improves their predictive capacity due to the balanced distribution.

6. Modeling

Embeddings in the world of natural language processing (NLP) represent words as vectors of real numbers, meaning that text is transformed into a point within a low-dimensional vector space [13]. Subsequently, the models can learn and capture specific features of the elements, which in turn identify semantic relationships, facilitating the comparison of other words with similar meanings.

The embedding models used for the representation of product comments are:

- **Word2Vec**: In this model, vectors result from an unsupervised learning process, where neural networks predict a word based on its contextual terms [31]. The Python library Gensim provides the `word2vec-google-news-300` model as a resource.
- **GloVe**: Words in comparable situations have a semantic relationship, and through a co-occurrence matrix, the links between these words can be inferred [32]. The embeddings are generated using the spaCy library through the `es_core_news_md` model.
- **MPNet** uses transformer-based language models for its pretraining, applying token masking and permutation to enhance the contextual understanding of words in the corpus [33]. Hugging Face provides the sentence transformer model `all-mpnet-base-v2`.
- **ADA-002** uses neural networks to convert text and code into vector representations, embedding them in a high-dimensional space [5]. The OpenAI library provides accessibility and simplicity for using the `text-embedding- ada-002` model.

For the vectorization of opinions in each ranking, a cleaning process is performed, primarily removing unnecessary characters and converting the embeddings into a suitable matrix format, without applying stopword removal or lemmatization processes, as these can affect the context and meaning of the phrases for vectorization.

Table 3
Comparison of embedding models by release year, dimensionality, and key characteristics

Model	Year	Dimensions	Type	Training Approach
Word2Vec	2013	100	Word embedding	Shallow neural network
GloVe	2014	300	Word embedding	Global co-occurrence statistics
MPNet	2020	768	Sentence embedding	Transformer (BERT-like)
ADA-002	2022	1,536	Sentence embedding	Contrastive learning (LLM-based)

Table 3 presents a comparison of the number of dimensions of the different embedding models, and various key characteristics. While Word2Vec is deemed suitable for basic semantic similarity tasks, GloVe is better suited for word analogy tasks, and MPNet in sentence-level retrieval and similarity. ADA-002, being LLM-based, is suitable for cross-domain semantic search and generalization [34].

Four classification models are employed for our experimental evaluation:

- **Logistic Regression** [35] is a supervised model whose purpose is to develop a binary classification. It converts the output of a linear model into a probability that can be used for categorization through the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-X}} \tag{1}$$

  This better captures the subtleties of the data, such as non-linear correlations, while reducing overfitting, which can lead to incorrect predictions [36]. To extend binary logistic regression to multiclass classification, the One-vs-Rest (OvR) approach was adopted, training one classifier per class to distinguish it from the remaining classes, with the final prediction selected based on the highest classifier confidence.
- **Support Vector Machines**: A popular classification model in machine learning due to its balanced predictive performance. This model performs an implicit mapping of variables into a high-dimensional feature space using a kernel function. This function returns the inner product $(\phi(x), \phi(x'))$ between two data points x, x' in the feature space [37]:

$$k(x, x') = (\phi(x), \phi(x')) \tag{2}$$

- **Random Forest**: It is a supervised modeling technique that uses multiple decision trees on a dataset. The results obtained are combined (typically through averaging or vote counting) to produce a single, more robust model compared to the results from each individual tree [38].
- **Decision Trees**: A decision tree is a classifier that expresses how a space of instances can be partitioned. Each internal node divides the instances into two or more subspaces based on a discrete function of the input attribute values and ranges when the attribute is numerical. The result for each leaf is a probability vector indicating the likelihood that the target attribute will have a specific value [39].

7. Evaluation

The evaluation is carried out in two steps: (i) vector embeddings using the different embedding models are generated from each *review text*, and (ii) the four classification models are trained to predict the *review score* using five-fold cross-validation.

The first metric that we report is *accuracy*, which is calculated as the sum of true positive (TP) and true negative (TN) predictions divided by the total number of data points (P + N) [40]. Accuracy is relevant in e-commerce sentiment analysis because it measures how often the model correctly classifies customer sentiments overall, ensuring reliable insights for improving products, services, and customer

experiences while minimizing potentially costly misclassifications. In Table 4, it is observed that the vectors generated by ADA-002 outperform the others, achieving the highest values with the Random Forest model (65.4%).

Table 4
Accuracy of models with each type of embedding

Embedding	Logistic Regression	Support Vector Machine	Random Forest	Decision Tree
Word2Vec	0.274400	0.356400	0.342000	0.250400
GloVe	0.501200	0.540800	0.560000	0.466400
MPNet	0.498800	0.566800	0.552400	0.476000
ADA-002	0.585200	0.648800	0.654400	0.554400

Subsequently, we present the results for *precision*, which measures the proportion of positive predictions made by the model that are actually correct [41]. Precision is a useful metric in e-commerce sentiment analysis as it quantifies the reliability of positive sentiment classifications. In Table 5, it is again observed that the embeddings generated using the ADA-002 model outperform the others and continue to achieve the highest values with the Random Forest model (65.1%).

Table 5
Precision of the models with each type of embedding

Embedding	Logistic Regression	Support Vector Machine	Random Forest	Decision Tree
Word2Vec	0.263057	0.349937	0.333715	0.250379
GloVe	0.501615	0.541324	0.558672	0.466969
MPNet	0.497026	0.565060	0.551112	0.477081
ADA-002	0.585659	0.644760	0.651162	0.554572

The next classification metric we consider is *specificity*, which measures the model's ability to correctly identify true negatives, assessing how well the model avoids false positives. The metric is calculated as the ratio of true negatives (TN) to the sum of true negatives (TN) and false positives (FP) [42]. In e-commerce sentiment analysis, specificity is important to accurately identify negative customer sentiments, minimizing false positives that could misclassify satisfied customers as dissatisfied. In Table 6, it can be observed that the embeddings generated using the ADA-002 model continue to outperform the others, achieving the highest values with the support vector machines model (91.2%) and closely followed by the random forest model, with a value of approximately 91.3%.

Table 6
Specificity of the models with each type of embedding

Embedding	Logistic Regression	Support Vector Machine	Random Forest	Decision Tree
Word2Vec	0.818612	0.839139	0.835440	0.812622
GloVe	0.875285	0.885169	0.889941	0.866551
MPNet	0.874669	0.891636	0.887990	0.868979
ADA-002	0.896324	0.912144	0.913496	0.888608

The last classification metric is the F1-score, which combines precision and recall into a single value, providing a balance between both, which is useful for imbalanced datasets [43]. In Table 7, it is identified that the vectors created with the ADA-002 model show the highest performance, with the highest values achieved using the Support Vector Machines model (64.5%), followed by the Random Forest model with a value of approximately 64.9%.

These results allow us to determine that vectorization with ADA-002 provides more information

Table 7
F1 Score of the models with each type of embedding

Embedding	Logistic Regression	Support Vector Machine	Random Forest	Decision Tree
Word2Vec	0.263962	0.341095	0.333619	0.250053
GloVe	0.501254	0.540880	0.557727	0.466014
MPNet	0.497698	0.564919	0.549551	0.476036
ADA-002	0.585261	0.645893	0.649979	0.554120

to the models for making predictions. One of the models with the best performance in the metrics is Support Vector Machines, primarily in balancing precision and recall, as seen in the F1 score of 64.5%.

7.1. An Example Run

Through vectorization, a methodology for natural language interpretation can be achieved. In this case, the use of embeddings facilitates the assignment of a rating to the comments made by customers. This rating ranges from 1 to 5 and provides feedback to the seller on the quality and service of their products. In Table 8, seven random comments are assigned and then vectorized using the ADA-002 embedding model. The Support Vector Machine model, which performs best during the modeling phase, is used to make the prediction. Negative comments are assigned a rating of 1 or 2, while reviews with a positive intention have ratings above 3.

Table 8
Rating new reviews with ADA-002

Comment	Vector	Prediction
The product is very good	[-0.00019131091539748013, 0.003384261624887585...	4
The headphones are of low quality	[-0.002562529407441616, 0.00621490553021431, -...	2
horrible product that only worked for 3 days	[-0.03443586453795433, 0.012141860090196133, 0...	1
I don't know what to think	[0.002560935216024518, -0.011984720826148987, ...	1
I want my money back	[-0.032338231801986694, -0.01803927682340145, ...	1
The keyboard is bad but the screen is perfect	[-0.012790611945092678, -0.000450057239504531,...	3
I love the mouse and it has great functionality...	[-0.02390960231423378, -0.00342733645811677, 0...	5

8. Conclusions

The study evaluated the effectiveness of Word2Vec, GloVe, MPNet, and ADA-002 embeddings for sentiment analysis on MercadoLibre product reviews, revealing ADA-002 as the top performer across all metrics, including accuracy, precision, specificity, and F1-score. Its high-dimensional, context-aware architecture enabled superior semantic understanding, particularly when paired with Support Vector Machines, which achieved an F1-score of 64.5% and specificity of 91.2%. These results highlight the advantages of transformer-based embeddings over traditional static methods like Word2Vec and GloVe, demonstrating their ability to capture nuanced sentiments in user reviews, where phrasing and context significantly influence meaning. However, ADA-002's computational intensity and proprietary nature

may raise concerns about scalability, reproducibility, and bias. Additionally, its reliance on API access potentially introduces latency and cost barriers.

The findings have immediate practical applications for e-commerce platforms, such as automated review moderation, enhanced recommendation systems, and vendor performance analytics. By deploying our approach, businesses can automatically detect negative sentiments for timely customer service responses, prioritize well-reviewed products in recommendations, and identify recurring product issues. By balancing imbalanced review data we ensure that the models generalize well while maintaining data integrity. This approach provides a scalable framework for sentiment analysis that can be adapted to other languages and domains.

Future research should explore multilingual extensions, domain-specific fine-tuning of embeddings, and real-time processing for live customer feedback. While ADA-002's performance sets a high benchmark, further investigation into the relationship between embedding dimensionality and model efficacy could yield fruitful results. This study underscores the transformative potential of advanced NLP techniques in e-commerce, offering actionable insights for businesses while paving the way for more sophisticated sentiment analysis tools in an increasingly digital marketplace.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o to perform grammar and spelling checks. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. Tsirmpas, I. Gkionis, G. T. Papadopoulos, I. Mademlis, Neural natural language processing for long texts: A survey on classification and summarization, Engineering Applications of Artificial Intelligence 133 (2024) 108231.

[2] A. Ramesh Kashyap, Y. Yang, M.-Y. Kan, Scientific document processing: challenges for modern learning methods, International Journal on Digital Libraries 24 (2023) 283–309.

[3] D. Ofori, Gpt-3 vs other text embeddings techniques for text classification: A performance evaluation, Medium (2023).

[4] Nimbleway, Why sentiment analysis is the missing link in your retail/e-commerce voc strategy (2025). URL: https://www.nimbleway.com/blog/why-sentiment-analysis-matters-in-retail.

[5] A. Neelakantan, T. Xu, R. Puri, A. Radford, J. M. Han, J. Tworek, Q. Yuan, N. Tezak, J. W. Kim, C. Hallacy, et al., Text and code embeddings by contrastive pre-training, arXiv preprint arXiv:2201.10005 (2022).

[6] S. Filipetto, Vender en mercadolibre, Papeles de trabajo: La revista electrónica del IDAES 17 (2023).

[7] L. Elliott, Latin America's e-commerce king says MercadoLibre has huge room for growth, https://www.reuters.com/technology/latin-americas-e-commerce-king-says-mercadolibre-has-huge-room-growth-2024-09-09/, 2025. Reuters, published 2025-05-27. Accessed 2025-09-14.

[8] R. Wirth, J. Hipp, Crisp-dm: Towards a standard process model for data mining, in: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, volume 1, Manchester, 2000, pp. 29–39.

[9] P. L. Rodriguez, A. Spirling, Word embeddings: What works, what doesn't, and how to tell the difference for applied research, The Journal of Politics 84 (2022) 101–115. doi:10.1086/715162.

[10] M. Xu, Understanding Graph Embedding Methods and Their Applications, volume 63, 2021. doi:10.1137/20M1386062.

[11] S. S. Birunda, R. K. Devi, A Review on Word Embedding Techniques for Text Classification, 2021, pp. 267–281. doi:10.1007/978-981-15-9651-3_23.

[12] M. Grohe, word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data, in: Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, ACM, 2020, pp. 1–16. doi:10.1145/3375395.3387641.
[13] H. W. Chung, T. Févry, H. Tsai, M. Johnson, S. Ruder, Rethinking embedding coupling in pre-trained language models, volume 1, 1 ed., 2020.
[14] S. J. Johnson, M. R. Murty, I. Navakanth, A detailed review on word embedding techniques with emphasis on word2vec, Multimedia Tools and Applications 83 (2023) 37979–38007. doi:10.1007/s11042-023-17007-z.
[15] N. Badri, F. Kboubi, A. H. Chaibi, Combining fasttext and glove word embedding for offensive and hate speech text detection, Procedia Computer Science 207 (2022) 769–778. doi:10.1016/j.procs.2022.09.132.
[16] L. Cao, R. Zeng, S. Peng, A. Yang, J. Niu, S. Yu, Textual emotion classification using mpnet and cascading broad learning, Neural Networks 179 (2024) 106582. doi:10.1016/j.neunet.2024.106582.
[17] M. A. Khder, Web scraping or web crawling: State of art, techniques, approaches and application, volume 13, 3 ed., 2021.
[18] M. Dogucu, M. Çetinkaya Rundel, Web scraping in the statistics and data science curriculum: Challenges and opportunities, Journal of Statistics and Data Science Education 29 (2021) S112–S122. doi:10.1080/10691898.2020.1787116.
[19] S. Lunn, J. Zhu, M. Ross, Utilizing web scraping and natural language processing to better inform pedagogical practice, in: 2020 IEEE Frontiers in Education Conference (FIE), IEEE, 2020, pp. 1–9. doi:10.1109/FIE44824.2020.9274270.
[20] V. Krotov, L. Johnson, L. Silva, Legality and ethics of web scraping, Communications of the Association for Information Systems 47 (2020) 539–563. doi:10.17705/1CAIS.04724.
[21] R. Brewer, B. Westlake, T. Hart, O. Arauza, The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection, Springer International Publishing, 2021, pp. 435–456. doi:10.1007/978-3-030-74837-1_22.
[22] C. C. de Comercio Electronico, Medición de indicadores -tendencia de la oferta de bienes y servicios en línea-, 2019.
[23] M. P. Perdomo, Percepción de confianza que genera a sus clientes el modelo de negocio de mercado libre, Universidad de La Salle (2018).
[24] mercadolibre, https://www.mercadolibre.com.co/, 2025.
[25] J. O. C. Casas, C. D. R. Castillos, Percepción y preferencia de compra de los clientes por medio de la plataforma mercado libre.com, Universidad Cooperativa de Colombia (2021).
[26] O. Aydin, R web scraping quick start guide techniques and tools to crawl and scrape data from websites, 2018. URL: www.packtpub.com.
[27] J. A. Sánchez, L. A. Montoya, La confianza como elemento fundamental en las compras a través de canales de comercio electrónico. caso de los consumidores en antioquia (colombia), Innovar 27 (2017) 11–22. doi:10.15446/innovar.v27n64.62365.
[28] C. D. Manning, An introduction to information retrieval, 2009.
[29] F. P. Sourd, Xml, json y el intercambio de información, ACUNAH 18 (2022).
[30] nltk, https://www.nltk.org/, 2025.
[31] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space (2013).
[32] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.
[33] K. Song, X. Tan, T. Qin, J. Lu, T.-Y. Liu, Mpnet: Masked and permuted pre-training for language understanding, Nanjing University of Science and Technology (2020).
[34] R. Patil, S. Boit, V. Gudivada, J. Nandigam, A survey of text representation and embedding techniques in nlp, IEEE Access 11 (2023) 36120–36146.

[35] D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2013. doi:10.1002/9781118548387.

[36] M. P. LaValley, Logistic regression, Circulation 117 (2008) 2395–2399. URL: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.106.682658. doi:10.1161/CIRCULATIONAHA.106.682658.

[37] D. Meyer, Support vector machines, R News 1 (2001) 23–26.

[38] J. J. E. Zúñiga, Aplicación de algoritmos random forest y xgboost en una base de solicitudes de tarjetas de crédito, Ingeniería Investigación y Tecnología 21 (2020) 1–16. doi:10.22201/fi.25940732e.2020.21.3.022.

[39] J. Quinlan, Simplifying decision trees, International Journal of Man-Machine Studies 27 (1987) 221–234. doi:10.1016/S0020-7373(87)80053-6.

[40] Željko Đ. Vujovic, Classification model evaluation metrics, International Journal of Advanced Computer Science and Applications 12 (2021). doi:10.14569/IJACSA.2021.0120670.

[41] P. St-Aubin, B. Agard, Precision and reliability of forecasts performance metrics, Forecasting 4 (2022) 882–903. doi:10.3390/forecast4040048.

[42] A. J. R. Villegas, M. Romero, N. Serna, Risk adjustment revisited using machine learning techniques, Documentos CEDE (2017).

[43] R. Yacouby, D. Axman, Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models, in: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, Association for Computational Linguistics, 2020, pp. 79–91. doi:10.18653/v1/2020.eval4nlp-1.9.