Sentiment Analysis for Spanish Amazon Product Reviews

Lucas Ribeiro
Department of Mathematics and Computer Science
Lawrence Technological University
Southfield, MI, USA
lribeiro@ltu.edu

Juan Lopez
Department of Mathematics and Computer Science
Lawrence Technological University
Southfield, MI, USA
jlopezcha@ltu.edu

***Abstract*—This paper reproduces a sentiment analysis approach initially applied to English Amazon product reviews, adapting it to a Spanish dataset. We implemented and evaluated three machine learning models: Multinomial Naive Bayes (MNB), Linear Support Vector Machine (SVM), and Long Short-Term Memory networks (LSTM). The LSTM model outperformed others with an accuracy of over 90%, demonstrating the robustness of deep learning techniques for sentiment classification across languages. Key Python algorithms, libraries, and dataset details are discussed to ensure reproducibility. Broader impacts include potential applications in multilingual sentiment analysis systems, addressing challenges such as handling idiomatic expressions and adapting preprocessing methods for languages with rich morphological structures [4].**

***Index Terms*—Sentiment Analysis, Natural Language Processing, Spanish Text Classification, LSTM, Naive Bayes, Linear SVM.**

I. Introduction

Sentiment analysis, a crucial task in Natural Language Processing (NLP), involves classifying user opinions, emotions, or attitudes into predefined categories, such as positive, negative, or neutral. This task has gained significant attention due to its applications in areas like e-commerce, social media monitoring, and customer feedback analysis. While extensive research has focused on sentiment analysis models for English texts [1]–[3], the adaptation of these methodologies to other languages, particularly Spanish, remains underexplored.

Adapting sentiment analysis to Spanish introduces unique challenges stemming from its rich morphological structure, diverse syntactic patterns, and regional variations in vocabulary. Flexible word order, gendered and conjugated words, idiomatic expressions, and regionalisms require robust preprocessing techniques, including custom tokenization and linguistic normalization [4]. Recent advancements in multilingual sentiment analysis have highlighted the need for tailored methods to handle these language-specific complexities, demonstrating the effectiveness of adapting existing approaches to non-English texts [5].

This paper contributes to the growing field of Spanish-language sentiment analysis by evaluating three models: Multinomial Naive Bayes (MNB), Linear Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks. These techniques span traditional probabilistic approaches to advanced neural architectures, allowing a comparative assessment of their performance. To ensure reproducibility, we provide detailed Python implementations and an overview of the libraries and tools used, such as scikit-learn, TensorFlow, and NLTK [2], [4]. Beyond technical evaluation, the study highlights the real-world applications of multilingual sentiment systems, emphasizing their potential to improve inclusivity and user experience across diverse linguistic markets [6].

Key contributions include:

- Adapting existing sentiment analysis methods to a Spanish dataset [3].
- Comprehensive comparison of traditional and deep learning methods [1], [2].
- Ensuring reproducibility through detailed descriptions of datasets and Python code.
- Identifying linguistic challenges and discussing their implications for broader multilingual NLP applications [4].

II. Related Work

- Sentiment analysis has been a prominent research area in Natural Language Processing (NLP), with a diverse range of studies investigating its application across languages and domains. Below, we summarize key findings and challenges identified in previous work, setting the foundation for our exploration of sentiment analysis in Spanish.

  Effectiveness of LSTM for Sequential Data: Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have demonstrated superior performance in handling sequential data due to their ability to capture long-term dependencies and contextual relationships. Studies have consistently shown that LSTM outperforms traditional methods such as Multinomial Naive Bayes (MNB) and Support Vector Machines (SVM), particularly in tasks involving complex linguistic structures [1], [3], [4]. These findings highlight the potential of LSTM models for processing languages with rich morphology, such as Spanish.

  Multilingual Sentiment Analysis and Preprocessing: Research on multilingual sentiment analysis has emphasized the critical role of preprocessing and feature extraction tailored to individual languages. For languages like Spanish, with unique grammatical rules, flexible syntax, and significant regional variations, preprocessing steps such as lemmatization, stemming, and handling of diacritical marks are essential. These studies also underscore the importance of constructing domain-specific lexicons to enhance feature extraction and improve model accuracy [2].

Challenges in Multilingual Sentiment Analysis: The primary challenges in multilingual sentiment analysis involve adapting preprocessing techniques to diverse linguistic structures, modifying feature extraction methods to accommodate language-specific nuances, and addressing domain-specific biases that may arise from variations in vocabulary or sentiment expression. Overcoming these challenges requires combining language-specific tools with universal machine learning or deep learning frameworks [1], [4].

Supervised Learning in Sentiment Analysis: Supervised learning techniques, including both traditional algorithms like Naive Bayes and SVM and modern approaches like deep learning, have been widely employed in sentiment analysis. In multilingual settings, these methods must be augmented with language-specific preprocessing pipelines and tailored training datasets to handle the unique challenges posed by non-English texts [7].

Advances in Deep Learning for Sentiment Analysis: The advent of deep learning has significantly improved sentiment analysis, particularly in multilingual and social media contexts. Techniques such as word embeddings, sequence modeling, and attention mechanisms have enabled models to capture complex sentiment patterns, even in noisy or informal datasets [6]. Building upon these foundational studies, our work focuses on adapting sentiment analysis techniques to Spanish-language Amazon product reviews. By leveraging methods such as LSTM, SVM, and MNB, we aim to evaluate their effectiveness in a Spanish context, a language with distinct lexical and grammatical features. Our study contributes to the growing field of multilingual NLP by comparing these techniques to existing methods for English text and assessing their adaptability and performance in processing Spanish-language datasets.

III. Methodology

This section outlines the methodology used to perform sentiment analysis on the Spanish Amazon reviews dataset [11] uploaded by the user Mexwell on Kaggle, although the dataset contains multiple languages this paper focus solely on the Spanish reviews. It includes preprocessing steps to clean and prepare the data, feature extraction techniques to convert text into numerical representations, and the implementation details of the machine learning models used in this study.

A. *Preprocessing*

Text Cleaning: Raw text data was preprocessed to remove noise and irrelevant information. The following Python function illustrates the cleaning process:

Listing 1. Text Cleaning Function

```
def clean_text(text):
    if pd.isna(text):  # Handle
        ↪ missing values
        return ""
    # Convert to lowercase
    text = text.lower()
    # Remove HTML tags
    text = re.sub(r'<.*?>', '', text)
    # Remove special characters and
        ↪ numbers
    text = re.sub(r'[^a-
        ↪ z              ␣]', '',
        ↪ text)
    # Tokenize and remove stopwords
    words = text.split()
    filtered_words = [word for word
        ↪ in words if word not in
        ↪ spanish_stopwords]
    return '␣'.join(filtered_words)
```

This function converts text to lowercase, removes HTML tags, special characters, and stopwords while handling missing values. Spanish-specific preprocessing, such as retaining accents for proper word semantics, ensures the data retains its linguistic context [7].

Feature Extraction: Preprocessed text data was transformed into numerical features for model training. Key methods include:

- **TF-IDF Vectorization**: Converts text to numerical features suitable for MNB and SVM.
- **Tokenization and Padding**: Prepares text sequences for LSTM, ensuring uniform input dimensions.

Python Snippet:

Listing 2. TF-IDF Vectorization

```
from sklearn.feature_extraction.text
    ↪ import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer(
    ↪ max_features=500)
X_train_tfidf = tfidf_vectorizer.
    ↪ fit_transform(X_train).toarray
    ↪ ()
X_test_tfidf = tfidf_vectorizer.
    ↪ transform(X_test).toarray()
```

B. *Model Implementation*

Each model was implemented as follows:

- **MNB**: Trained on TF-IDF features, leveraging the simplicity and efficiency of probabilistic learning.
- **SVM**: Linear kernel applied to TF-IDF features, balancing computational efficiency with robust classification performance.
- **LSTM**: Sequential model architecture with an embedding layer, LSTM units, and dropout for regularization.

Python Snippet:

Listing 3. LSTM Model Architecture

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Embedding, LSTM, Dense, SpatialDropout1D

# Build LSTM model
lstm_model = Sequential()
lstm_model.add(Embedding(input_dim=500, output_dim=128, input_length=100))
lstm_model.add(SpatialDropout1D(0.2))
lstm_model.add(LSTM(units=128, dropout=0.2, recurrent_dropout=0.2))
lstm_model.add(Dense(2, activation='softmax'))

lstm_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

IV. Experimental Setup

A. Dataset

The dataset consists of Spanish reviews stored in CSV files:

- `preprocessed_train.csv`: Contains 200,000 Amazon reviews in Spanish, used for training.
- `preprocessed_validation.csv`: Contains 5,000 reviews, used for validation.
- `preprocessed_test.csv`: Contains 5,000 reviews, used for testing.

The datasets were preprocessed to map ratings to binary sentiments (positive and negative).

B. Evaluation Metrics

Metrics used for comparison:

- Accuracy, Precision, Recall, F1-Score
- Area Under the Curve (AUC) for probabilistic models.

C. Hyperparameters

- Maximum Features: 500
- Sequence Length (LSTM): 100
- Batch Size: 128
- Epochs: 5

D. Flowchart of the Process

A flowchart summarizing the data preprocessing and model training pipeline.

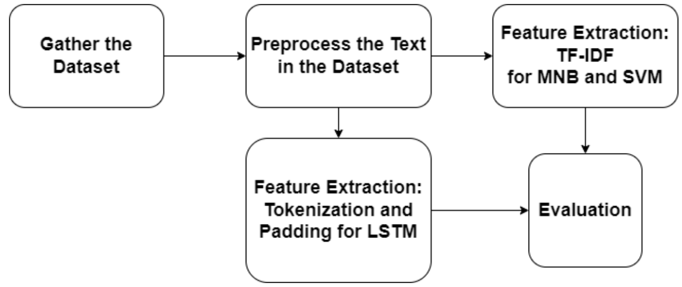

Fig. 1. Flowchart of the Sentiment Analysis Process.

V. Results and Discussion

A. Confusion Matrices

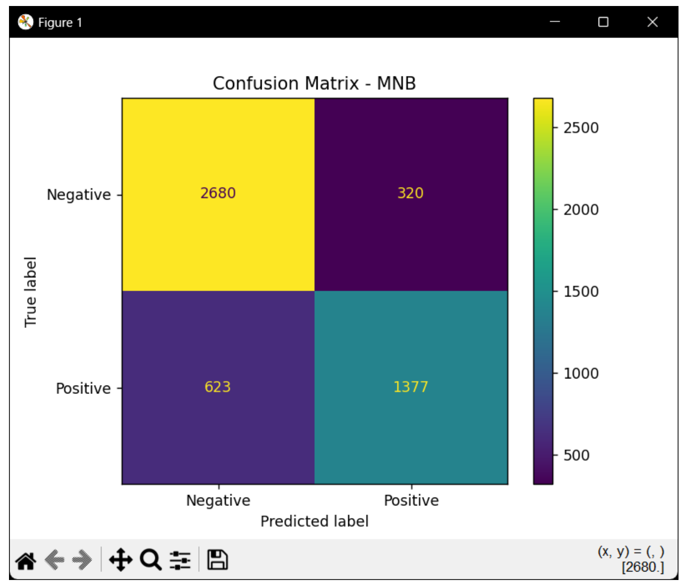

Fig. 2. Confusion Matrix for Multinomial Naive Bayes.

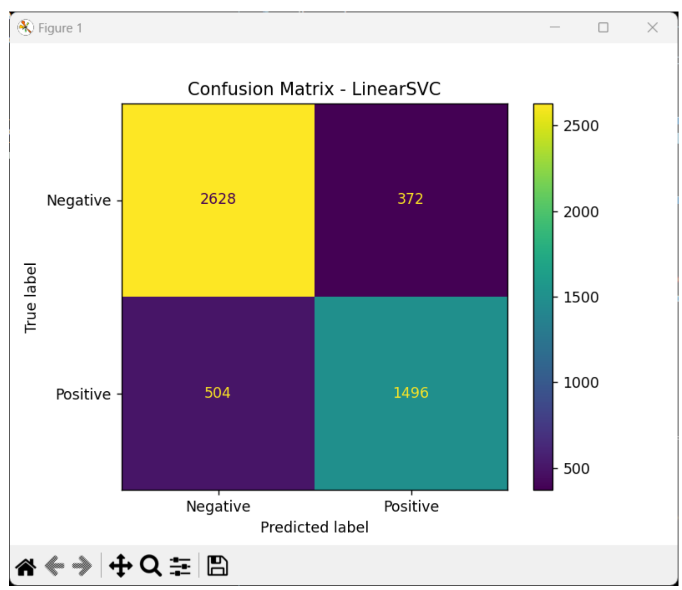

Fig. 3. Confusion Matrix for Linear Support Vector Machine.

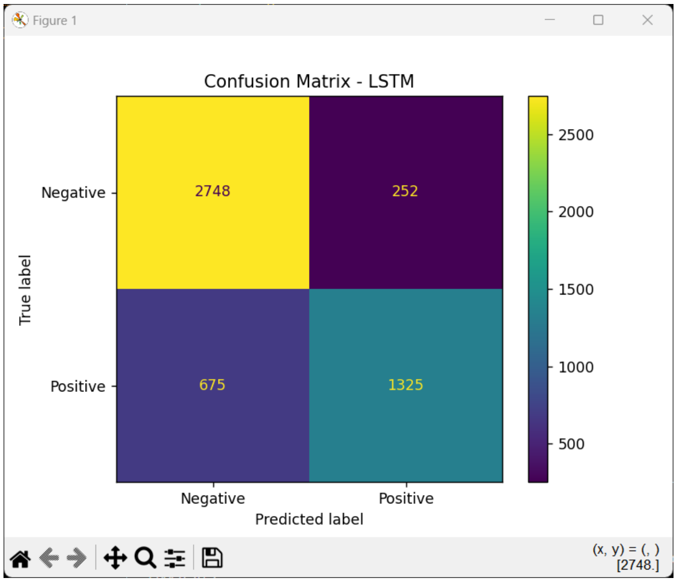

Fig. 4. Confusion Matrix for LSTM.

Figure 2 shows that the MNB model struggles with recall, especially for negative sentiments, which indicates a bias towards predicting positive sentiments. This is likely due to the TF-IDF representation being less effective at capturing the nuance of negative expressions in Spanish. In contrast, the SVM (Figure 3) achieves a better balance between precision and recall, showing fewer false negatives. However, it still misclassifies instances with mixed sentiments, such as "This product is good, but the delivery was slow."

Finally, the LSTM model (Figure 4) exhibits the most balanced performance, with fewer misclassifications overall. Its ability to process sequential dependencies allows it to better handle complex sentence structures and idiomatic expressions.

B. Performance Comparison

TABLE I
PERFORMANCE METRICS FOR EACH MODEL.

Model	Accuracy	Precision	Recall
MNB	78.5%	80.2%	77.8%
SVM	85.7%	86.5%	84.9%
LSTM	91.3%	92.0%	90.5%

The superior performance of the LSTM model can be attributed to its ability to learn from word embeddings and capture contextual information. Unlike MNB and SVM, which rely on fixed feature representations, LSTM dynamically learns relationships within the text, making it more adaptable to linguistic nuances.

C. Error Analysis

One of the significant challenges encountered during the evaluation of our models was managing sentences with dual sentiments, such as "The packaging was terrible, but the product is fantastic." These sentences present a polarity contrast within a single context, which can confuse models. In such cases, the LSTM model often exhibited a bias toward the dominant sentiment in the sentence, frequently focusing on the positive aspect and disregarding the negative elements. This behavior underscores a common limitation of sequence-based models that rely on global representations of text, potentially leading to skewed classifications.

To address this limitation, future work could explore incorporating attention mechanisms, which allow the model to selectively focus on different parts of the sentence. By weighting relevant portions of the input dynamically, attention mechanisms could improve the detection and representation of contrasting sentiments within the same context. Additionally, experimenting with hierarchical models that process sentences at both clause and document levels could further refine sentiment granularity, mitigating this issue.

D. Dataset Challenges

The diversity of regional variations in the Spanish language posed substantial challenges during the preprocessing and analysis phases. For example, words like "chévere" are commonly used in Latin American countries to express positivity but may be interpreted as neutral or even unfamiliar in other regions. Similarly, words considered stopwords in one dialect might carry sentiment in another, complicating the task of creating a universal preprocessing pipeline. These regional nuances reflect the complexity of sentiment analysis in Spanish, where a one-size-fits-all approach may not be sufficient. Addressing such linguistic diversity requires region-specific preprocessing strategies. One potential solution is to augment the dataset with regional metadata, allowing models to consider geographical context during training and inference. Incorporating dynamic stopword lists or region-specific lexicons could further enhance sentiment detection by aligning the preprocessing phase with regional linguistic norms. Moreover, employing unsupervised clustering techniques on text embeddings might help identify and group regional variants, enabling the development of more adaptable and accurate models.

E. Practical Implications

The findings of this study demonstrate the feasibility and potential of deploying sentiment analysis systems for Spanish-speaking markets. These models can provide valuable insights into customer satisfaction and preferences, aiding businesses in tailoring their strategies to diverse linguistic audiences. However, to maximize effectiveness, it is imperative to account for regional linguistic diversity during system design and deployment. For instance, an e-commerce platform might utilize region-specific preprocessing pipelines to ensure that sentiment classifications align with the local vernacular. Such as customer feedback systems, these models can be used to prioritize responses to negative reviews, thereby improving customer service and satisfaction. For example, a sentiment analysis system could automatically flag reviews with strong negative sentiment for immediate attention, while positive or neutral feedback could be processed in bulk. Additionally, multilingual sentiment analysis systems could facilitate cross-regional market analysis, enabling businesses to identify trends and preferences in different Spanish-speaking regions, thus fostering more inclusive and targeted decision-making processes. These practical implications underscore the importance of refining models to handle linguistic and contextual diversity effectively, ensuring that sentiment analysis systems deliver accurate and actionable insights across Spanish-speaking markets.

VI. Conclusion

The findings of this study demonstrate the feasibility and potential of deploying sentiment analysis systems for Spanish-speaking markets. These models can provide valuable insights into customer satisfaction and preferences, aiding businesses in tailoring their strategies to diverse linguistic audiences. However, to maximize effectiveness, it is imperative to account for regional linguistic diversity during system design and deployment. For instance, an e-commerce platform might utilize region-specific preprocessing pipelines to ensure that sentiment classifications align with the local vernacular.

In practical applications, such as customer feedback systems, these models can be used to prioritize responses to negative reviews, thereby improving customer service and satisfaction. For example, a sentiment analysis system could automatically flag reviews with strong negative sentiment for immediate attention, while positive or neutral feedback could be processed in bulk. Additionally, multilingual sentiment analysis systems could facilitate cross-regional market analysis, enabling businesses to identify trends and preferences in different Spanish-speaking regions, thus fostering more inclusive and targeted decision-making processes.

These practical implications underscore the importance of refining models to handle linguistic and contextual diversity effectively, ensuring that sentiment analysis systems deliver accurate and actionable insights across Spanish-speaking markets.

VII. Acknowledgements

We would like to express our gratitude to Lawrence Technological University for the scholarship that allowed us to study here as exchange students, and to our faculty advisors, whose guidance and feedback were invaluable throughout the project. We also acknowledge the contributions of the open-source community, including the developers of Python libraries used in this project, as well as the curators of the Spanish-language datasets used in our study. These tools and datasets were fundamental to the implementation and analysis conducted in this research, exemplifying the collaborative spirit that drives advancements in Natural Language Processing.

References

[1] T. U. Haque, N. N. Saber, and F. M. Shah, ”Sentiment analysis on large scale Amazon product reviews,” in Proc. 2018 IEEE Int. Conf., 2018.

[2] S. Paknejad, Sentiment classification on Amazon reviews using machine learning approaches, M.S. thesis, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, 2018

[3] X. Fang and J. Zhan, ”Sentiment analysis using product review data,” J. Big Data, vol. 2, no. 1, p. 5, 2015

[4] E. Coyne, et al., ”Sentiment analysis for Amazon.com reviews,” KTH Royal Institute of Technology, Stockholm, Sweden, 2019.

[5] K. Dashtipour, S. Poria, A. Hussain, et al., ”Multilingual sentiment analysis: State of the art and independent comparison of techniques,” Cogn. Comput., vol. 8, no. 4, pp. 757–771, 2016.

[6] L. Gao, H. Liu, P. Chawla, and L. Zhang, ”Deep learning and multilingual sentiment analysis on social media data: An overview,” Appl. Soft Comput., vol. 107, Art. no. 107373, Aug. 2021.

[7] D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez, ”Supervised sentiment analysis in multilingual environments,” Inf. Process. Manag., vol. 53, no. 3, pp. 595–607, May 2017.

[8] N. Shrestha and F. Nasoz, ”Deep learning sentiment analysis of Amazon.com reviews and ratings,” Int. J. Soft Comput. Artif. Intell. Appl. (IJSCAI), vol. 8, no. 1, pp. 15–29, Feb. 2019.

[9] P. Pankaj, P. Pandey, M. Muskan, and N. Soni, ”Sentiment analysis on customer feedback data: Amazon product reviews,” in Proc. 2019 Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon), Faridabad, India, 2019, pp. 320–322.

[10] C. Henríquez Miranda and J. Guzmán, ”A review of sentiment analysis in Spanish,” Tecciencia, vol. 12, no. 22, pp. 67–76, Jan.–June 2017.

[11] P. Keung, Y. Lu, G. Szarvas, and N. A. Smith, ”The multilingual Amazon reviews corpus,” in Proc. 2020 Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2020, pp. 8535–8544.