Multi-Dimensional Classification System using Pre-Trained Transformer Models for Multilingual Text Analysis

Emmanuel Arturo Torres-Santana[1,†], Manuel Enrique Balan-Euan[1,*,†]

[1]Instituto Tecnológico de Mérida, Computer Systems Engineering, Mérida, Yucatán, México

Abstract

This document presents a multi-dimensional classification system that integrates three pre-trained Transformer models for comprehensive text analysis. The system combines zero-shot classification using XLM-RoBERTa, sentiment analysis with multilingual BERT, and named entity recognition with Spanish BERT. The proposed architecture enables simultaneous classification across multiple dimensions, providing a robust solution for natural language processing in multilingual applications. Experimental results demonstrate the effectiveness of the ensemble approach for complex classification tasks.

Keywords

Natural Language Processing, Transformer Models, Multi-Dimensional Classification, Sentiment Analysis, Entity Recognition, Zero-Shot Classification

1. Introduction

Natural Language Processing (NLP) has experienced significant advances with the introduction of Transformer architectures and pre-trained models. These models have demonstrated exceptional capabilities across a wide range of tasks, from text understanding to language generation. However, many real-world applications require multidimensional text analysis that goes beyond a single classification task [1, 2, 3, 4, 5].

This work presents an integrated system that combines three specialized Transformer models to provide comprehensive text analysis across multiple dimensions: zero-shot thematic classification, sentiment analysis, and named entity recognition. The main motivation is to create a unified solution that can extract rich and multifaceted information from texts in different languages, particularly focused on Spanish and other resource-limited languages [6, 7, 8, 9].

The proposed system utilizes three main components: XLM-RoBERTa for zero-shot classification, multilingual BERT for sentiment analysis, and specialized Spanish BERT for named entity recognition. This combination allows simultaneous addressing of different aspects of textual analysis, providing a more complete understanding of the analyzed content.

The main challenges include effective integration of multiple models, computational complexity management, and performance optimization for real-time applications. Our approach addresses these challenges through a modular architecture that enables parallel execution and intelligent result aggregation.

Experimental results demonstrate that the ensemble system significantly outperforms individual models in complex classification tasks while maintaining acceptable computational efficiency for practical applications.

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

†These authors contributed equally.

LE21081354@merida.tecnm.mx (E. A. Torres-Santana); halomanue20@gmail.com (M. E. Balan-Euan)

https://github.com/SystemTrabu (E. A. Torres-Santana); https://github.com/ManuBalan03/ (M. E. Balan-Euan)

2. Related Work

Text classification has evolved from traditional methods based on manual features to deep approaches based on neural networks. Transformer models, introduced by Vaswani et al., revolutionized the NLP field by enabling parallel processing and capturing long-range dependencies more effectively than recurrent architectures.

BERT (Bidirectional Encoder Representations from Transformers) marked a milestone by introducing bidirectional pre-training, allowing models to capture context from both left and right sides of each token. Its multilingual variants, such as mBERT and XLM-RoBERTa, extended these capabilities to multiple languages, demonstrating effective cross-lingual transfer.

Named entity recognition (NER) has traditionally been approached as a sequential labeling task. BERT models have shown excellent performance in NER by combining deep contextual representations with specialized classification layers. For Spanish, models like BETO and specific variants have demonstrated significant improvements over previous methods.

Zero-shot classification represents an emerging paradigm where models can classify text into categories not seen during training. This is particularly valuable for applications where defining categories a priori is difficult or where categories change dynamically.

Ensemble systems that combine multiple specialized models have shown consistent advantages over individual models, albeit with the cost of greater computational complexity. Our work contributes to this line of research by proposing a specific architecture for integrating specialized Transformer models.

3. Tourism Context

This study is based on the **Rest-Mex 2025** corpus [10, 11], a large-scale dataset of tourist reviews focused on the most iconic and visited towns across Mexico. The dataset consists of 208,051 annotated entries, each representing a tourist's opinion and metadata collected from multiple sources. It was released as part of the "Sentiment Analysis Magical Mexican Towns" research initiative, and is intended exclusively for academic and research purposes.

Each review includes a title, a textual review, and three key labels:

- **Polarity**: A sentiment score from 1 (very dissatisfied) to 5 (very satisfied).
- **Type**: The category of place described, labeled as `Hotel`, `Restaurant`, or `Attractive`.
- **Geographic Location**: The name of the town and its corresponding region (state) in Mexico.

The corpus spans opinions from 40 carefully selected Mexican towns, such as Tulum, Isla Mujeres, San Cristóbal de las Casas, and Valladolid — places known for their cultural, historical, or ecological significance. These locations are distributed across 40 different states of Mexico, reflecting a wide geographic and touristic diversity. Tulum alone accounts for over 45,000 reviews, while towns like Tapalpa and Real de Catorce have under 1,000 reviews each, illustrating a natural class imbalance in the distribution of the data.

From a classification perspective, this dataset presents a multi-label challenge that involves:

1. Assigning a sentiment polarity (ordinal classification).
2. Identifying the type of business or site (nominal classification).
3. Predicting the corresponding municipality and state (geographic classification).

Such a setting closely simulates real-world tourism dynamics where both subjective experiences (e.g., satisfaction) and structured information (e.g., location and service type) coexist. The dataset not only enables experiments in multilingual NLP and sentiment analysis but also provides a practical foundation for regional tourism recommendation systems or public policy insights.

All preprocessing and model training in this research strictly adhere to the terms of academic use specified by the Rest-Mex 2025 initiative.

4. Natural Language Processing

To ensure high-quality input for training our classification models, a comprehensive text preprocessing pipeline was implemented. The process involved several Natural Language Processing (NLP) libraries and custom strategies to clean and normalize Spanish-language tourist reviews.

Initially, each review was transformed to lowercase and stripped of trailing whitespace. We used regular expressions to replace date patterns and long numerical sequences with placeholder characters to prevent misleading token frequency patterns. Then, we applied two normalization strategies from the `spanlp` library: `NumbersToVowelsInLowerCase` and `NumbersToConsonantsInLowerCase`. These strategies replaced digits with semantically plausible letters, which helped maintain syntactic structure without numerical noise.

Accents and diacritics were removed using Unicode normalization (`unicodedata`), followed by filtering all non-alphabetical characters. Tokenization was performed using the NLTK library, and Spanish stopwords were removed based on an extended list. After this step, we applied lemmatization using the `spaCy` model `es_core_news_sm`, which converts each token to its base form, improving the semantic consistency of inputs.

This preprocessing logic was encapsulated in a class named `Preprocesador`, which was used to transform each row in the dataset before being written to three FastText-compatible training files: one for sentiment polarity, one for business type (restaurant, hotel, or attraction), and one for geographic location (state and municipality). Only entries with valid and non-empty reviews and labels were included.

The resulting processed datasets were saved in plain text files following the FastText input format, with lines like:

```
__label__positivo excelente comida y servicio atencion rapida
__label__hotel lugar tranquilo y limpio en el centro
__label__yucatan-merida paseo cultural interesante y economico
```

These datasets were then used to train three independent classification models using the FastText library, achieving efficient training with n-gram features and dimensional embeddings. All these steps were crucial in preparing data that is both clean and semantically rich for downstream classification tasks.

This FastText-compatible structure allowed efficient supervised training using n-gram word representations and low-dimensional embeddings, while remaining computationally lightweight and suitable for large-scale experimentation.

5. Methodology

5.1. System Architecture

The proposed system consists of three main components that operate in a coordinated manner to provide multidimensional analysis of input text. The architecture was designed following principles of modularity and scalability.

5.1.1. Zero-Shot Classification Component

We utilize the `joeddav/xlm-roberta-large-xnli` model through the Hugging Face Transformers library. This model, based on XLM-RoBERTa, was trained on the XNLI (Cross-lingual Natural Language Inference) corpus and is capable of performing classification without prior examples in multiple languages.

Zero-shot classification is implemented through natural language inference hypothesis formulation. For each candidate category, a hypothesis of the type "This text is about [category]" is constructed and the probability of logical implication between the input text and the hypothesis is calculated.

5.1.2. Sentiment Analysis Component

Sentiment analysis is performed using the `nlptown/bert-base-multilingual-uncased-sentiment` model, which provides sentiment classification on a 5-point scale (1-5 stars). This model was specifically trained on multilingual reviews and demonstrates robustness across different textual domains.

5.1.3. Entity Recognition Component

For NER, we employ `mrm8488/bert-spanish-cased-finetuned-ner` with simple aggregation strategy. This model is specifically fine-tuned for Spanish and recognizes standard entities such as persons, organizations, locations, and other relevant categories.

5.2. Processing Pipeline

Algorithm 1 describes the complete processing flow:

Algorithm 1 Multi-Dimensional Classification Pipeline

Input: Input text T, candidate categories C
Initialize models: M_{zero}, M_{sent}, M_{ner}

// Parallel processing
$prob_{zero} \leftarrow M_{zero}(T, C)$ // Zero-shot classification
$sent_{score} \leftarrow M_{sent}(T)$ // Sentiment analysis
$entities \leftarrow M_{ner}(T)$ // Entity recognition

// Result aggregation
$result \leftarrow \{$
  $"classification" : prob_{zero},$
  $"sentiment" : sent_{score},$
  $"entities" : entities$
$\}$
Return $result$

5.3. Result Aggregation

The integration of results from the three models is performed through a unified data structure that preserves task-specific information while providing a coherent interface for downstream applications.

For global system evaluation, we define a composite metric that considers performance across all three dimensions:

$$Score_{global} = \alpha \cdot F1_{zero} + \beta \cdot F1_{sent} + \gamma \cdot F1_{ner} \tag{1}$$

where $\alpha + \beta + \gamma = 1$ and weights are adjusted according to the relative importance of each task in the specific application.

6. Implementation

6.1. Technical Configuration

The system was implemented in Python using the following main libraries:

- **Transformers** (v4.21.0): For loading and executing pre-trained models
- **PyTorch** (v1.12.0): Underlying deep learning framework
- **NumPy** (v1.23.0): Numerical operations and array manipulation
- **Pandas** (v1.4.3): Data processing and analysis

6.2. Model Initialization

Model loading was optimized to minimize initialization time and memory usage:

Listing 1: System Initialization

```python
from transformers import pipeline

# Zero-shot classifier
zero_shot_classifier = pipeline(
    "zero-shot-classification",
    model="joeddav/xlm-roberta-large-xnli"
)

# Sentiment analyzer
sentiment_classifier = pipeline(
    "sentiment-analysis",
    model="nlptown/bert-base-multilingual-uncased-sentiment"
)

# Named entity recognizer
ner = pipeline(
    "ner",
    model="mrm8488/bert-spanish-cased-finetuned-ner",
    aggregation_strategy="simple"
)
```

6.3. Performance Optimizations

To improve system efficiency, we implemented several optimizations:

1. **Batch processing**: Texts are processed in groups to leverage GPU parallelization
2. **Result caching**: Results from previously processed texts are stored to avoid re-computation
3. **Intelligent truncation**: Long texts are truncated while preserving semantically relevant information
4. **Model quantization**: Numerical precision reduction to accelerate inference

7. Experimental Evaluation

7.1. Evaluation Datasets

To evaluate system performance, we used multiple datasets covering different aspects of textual analysis:

- **XNLI-es**: Spanish subset of the XNLI corpus for zero-shot classification evaluation
- **TASS-2020**: Spanish sentiment analysis corpus from Twitter
- **CoNLL-2002 NER**: Standard dataset for Spanish entity recognition
- **MLDoc**: Multilingual document classification corpus

7.2. Evaluation Metrics

We evaluated performance using standard metrics for each task:

- **Classification**: Precision, Recall, macro and micro F1-score
- **Sentiment**: Accuracy, weighted F1, MAE for ordinal scores
- **NER**: F1 per entity, micro and macro F1, precision and recall

7.3. Results

Table 1 shows comparative results of the integrated system versus individual models:

Table 1
Comparative System Performance

Model/System	Zero-Shot F1	Sentiment Acc	NER F1	Global Score
XLM-RoBERTa alone	0.847	-	-	-
BERT Sentiment alone	-	0.792	-	-
BERT NER alone	-	-	0.891	-
Integrated System	0.851	0.798	0.894	0.848

7.4. Computational Performance Analysis

Computational performance analysis was conducted on an Amazon EC2 m6i.large instance (CPU-only) and reveals the following characteristics:

- **Inference time**: 1.2s average per text (500 tokens) on CPU
- **Memory usage**: 16GB RAM with all three models loaded
- **Throughput**: 230 texts/minute on EC2 m6i.large instance
- **Hardware specifications**: Intel Xeon processors, 32GB RAM, no GPU acceleration

8. Use Cases and Applications

8.1. Social Media Analysis

The system has been successfully applied to social media content analysis, where automatic topic classification, sentiment analysis, and entity extraction provide valuable insights for researchers and marketing analysts.

8.2. Corporate Document Processing

In enterprise environments, the system facilitates automatic document classification, relevant information extraction, and integrated customer feedback analysis.

8.3. Research Tools

Researchers in social sciences and digital humanities use the system for textual corpus analysis, enabling automated identification of thematic patterns, sentiment trends, and entities of interest.

9. Limitations and Future Work

9.1. Current Limitations

The system presents several limitations that should be considered:

- **Computational cost**: Simultaneous execution of three models requires significant computational resources
- **Language dependency**: Although multilingual, performance varies by specific language
- **Domain specificity**: Models may require fine-tuning for highly specialized domains
- **Latency**: Processing time may be prohibitive for ultra-low latency applications

9.2. Future Directions

Future work will focus on:

1. **Model optimization**: Implementation of distillation and pruning techniques to reduce size and accelerate inference
2. **Continual learning**: Development of mechanisms to update models with new data without complete retraining
3. **Modality integration**: System extension to process text, images, and audio in a unified manner
4. **Personalization**: Development of techniques to adapt the system to specific user needs

10. Conclusions

This work presents an integrated system for multi-dimensional text classification that effectively combines three specialized Transformer models. Experimental results demonstrate that the ensemble approach consistently outperforms individual models, providing richer and more reliable analysis of textual content.

The main contributions include: (1) a modular architecture for specialized model integration, (2) performance optimizations for practical applications, and (3) comprehensive evaluation across multiple datasets and metrics.

The system demonstrates practical applicability across various domains, from social media analysis to corporate document processing. Although limitations exist in terms of computational cost and domain specialization, future research directions offer promising paths to address these challenges.

Future research would benefit from exploring model compression techniques, end-to-end multi-task learning, and multi-modal extensions of the proposed approach.

Declaration on Generative AI

We declare that the present manuscript has been written entirely by the authors and that no generative artificial intelligence tools were used in its preparation, drafting, or editing.

References

[1] M. A. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-Gonzalez, D. Fajardo-Delgado, M. G. Sánchez, H. Pérez-Espinosa, J. Martínez-Miranda, R. Guerrero-Rodríguez, L. Bustio-Martínez, Ángel Díaz-Pacheco, Natural language processing applied to tourism research: A systematic review and future research directions, Journal of King Saud University - Computer and Information Sciences 34 (2022) 10125–10144. URL: https://www.sciencedirect.com/science/article/pii/S1319157822003615. doi:https://doi.org/10.1016/j.jksuci.2022.10.010.

[2] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancun case, seen from the usa, canada, and mexico, International Journal of Tourism Cities 10 (2023) 639–661. URL: http://dx.doi.org/10.1108/IJTC-09-2022-0223. doi:10.1108/ijtc-09-2022-0223.

[3] R. Guerrero-Rodríguez, M. A. Álvarez-Carmona, R. Aranda, et al., Big data analytics of online news to explore destination image using a comprehensive deep-learning approach: a case from mexico, Information Technology & Tourism 26 (2024) 147–182. URL: https://doi.org/10.1007/s40558-023-00278-5. doi:10.1007/s40558-023-00278-5.

[4] A. Diaz-Pacheco, M. A. Álvarez-Carmona, R. Guerrero-Rodríguez, L. A. C. Chávez, A. Y. Rodríguez-González, J. P. Ramírez-Silva, R. Aranda, Artificial intelligence methods to support the research of destination image in tourism. a systematic review, Journal of Experimental & Theoretical Artificial Intelligence 0 (2022) 1–31. doi:10.1080/0952813X.2022.2153276.

[5] Á. Díaz-Pacheco, R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, A. Y. Rodríguez-GonzÁlez, R. Aranda, A comprehensive deep learning approach for topic discovering and sentiment analysis of textual information in tourism, Journal of King Saud University - Computer and Information Sciences 35 (2023) 101746. URL: http://dx.doi.org/10.1016/j.jksuci.2023.101746. doi:10.1016/j.jksuci.2023.101746.
[6] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cárdenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism, Procesamiento del Lenguaje Natural 67 (2021). doi:https://doi.org/10.26342/2021-67-14.
[7] M. Á. Álvarez-Carmona, R. Aranda, R. Guerrero-Rodríguez, A. Y. Rodríguez-González, A. P. López-Monroy, A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one, Computación y Sistemas 26 (2022). doi:https://doi.org/10.13053/CyS-26-2-4055.
[8] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts, Procesamiento del Lenguaje Natural 69 (2022).
[9] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Muñis-Sánchez, A. P. Pastor-López, F. Sánchez-Vega, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, Procesamiento del Lenguaje Natural 71 (2023).
[10] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2025: Researching sentiment evaluation in text for mexican magical towns, volume 75, 2025.
[11] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

A. Online Resources

The source code and datasets used in this work are available at:

- GitHub Repository
- Hugging Face Models
- Datasets on Zenodo