

# Enhancing Transformer-Based Sentiment Analysis for the Rest-Mex 2025 Challenge: A Hybrid Strategy with Oversampling, Back-Translation, and Transformers

Muhammad Imran<sup>1,\*</sup>, Tayyab Rasheed<sup>2</sup> and Carlos Gómez-Rodríguez<sup>1</sup>

<sup>1</sup>Universidad de Coruña, CITIC Departamento de Ciencias de la Computación y Tecnologías de la Información, Campus de Elviña s/n, 15071, A Coruña, Spain

<sup>2</sup>Department of Computer Science, COMSATS University Islamabad, Pakistan

## Abstract

This paper presents a sentiment analysis framework for the Rest-Mex 2025 challenge, focused on Spanish-language reviews of Mexican Magical Towns. The task involves predicting sentiment polarity (1–5), classifying attraction type (Hotel, Restaurant, Attraction), and identifying the correct town from a list of 60. To address class imbalance, we propose a hybrid augmentation approach combining oversampling and back-translation using both structurally similar and dissimilar languages. Two transformer-based models roberta-base-bne and twitter-xlm-roberta-base are fine-tuned on the augmented datasets. The hybrid strategy, particularly with the multilingual model, achieved the best results, demonstrating improved performance and generalization across all subtasks. Our system achieved 4th place in the overall sentiment analysis track of the Rest-Mex 2025 shared task competing against 35 participating teams which demonstrates the robustness of our approach in sentiment classification across multiple subtasks.

## Keywords

Sentiment Analysis, TripAdvisor reviews, Transformers, NLP, Oversampling, Back-translation

## 1. Introduction

The sentiment analysis of the reviews provided by tourists plays a key role in the understanding of the public opinion in the tourism sector [1, 2] especially where tourism is the primary business. The analysis of Spanish-language reviews for Mexico’s Magical Towns (Pueblos Mágicos) introduces a unique set of linguistic and contextual challenges[3]. The reviews often contain informal tone, regional dialects, and culturally rooted expressions, which makes the task more complex than standard sentiment classification [4]. Unlike past editions [5, 6, 7], the Rest-Mex 2025 challenge at IberLef [8, 9] outlines a comprehensive framework that comprises three subtasks. The first is the prediction of sentiment polarity on a scale of 1 to 5. The second involves classification of the type of destination (either hotel, restaurant, or attraction). The third and final subtask is the identification of the corresponding Magical Town from a predefined list of 60 towns.

The field of sentiment analysis has seen widespread adoption of transformer-based architectures. Models such as BERT, BETO, and XLM-R have shown strong results on various multilingual and Spanish-specific datasets [10]. Prior to the advent of transformers, classical machine learning algorithms like SVMs and logistic regression were commonly used, often relying on handcrafted linguistic features and bag-of-words representations [11]. While these earlier approaches provided reasonable results, they lacked the deep contextual understanding required for handling nuanced and complex reviews.

The major limitation of existing models lies in the handling of class imbalance [12]. Sentiment labels such as “very negative” or “very positive” are often underrepresented, as are reviews for certain Magical Towns. The general-purpose nature of most pretrained models results in poor handling of

---

IberLEF 2025, September 2025, Zaragoza, Spain

\*Corresponding author.

✉ m.imran@udc.es (M. Imran); tayyabrasheed135@gmail.com (T. Rasheed); carlos.gomez@udc.es (C. Gómez-Rodríguez)

ORCID 0000-0002-4124-7929 (M. Imran); 0000-0003-0752-8812 (C. Gómez-Rodríguez)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

tourism-specific phrases. Furthermore, the joint modeling of sentiment, destination type, and town identification remains a non-trivial challenge due to the multitask nature of the problem.

To address the class imbalance and improve performance across all subtasks, we introduce three targeted strategies. The first involves the use of oversampling [13] for underrepresented classes. The second applies back-translation [14] using structurally similar and dissimilar languages to increase the diversity and volume of minority-class data. The third combines both techniques in a hybrid approach. For model training, we use the `roberta-base-bne` and `twitter-xlm-roberta-base` checkpoints, fine-tuned independently on datasets prepared with each strategy. Our solution improves the representation of low-frequency classes and enhances generalization across the three subtasks.

## 2. Related Work

Recent studies highlight the success of transformer architectures such as multilingual BERT, XLM-RoBERTa, and domain-specific models for sentiment and text classification. For instance, BETO has shown improved performance on Spanish sentiment tasks, while multilingual models allow for knowledge transfer across languages. Back-translation and data augmentation strategies have also gained attention for mitigating data imbalance and enhancing robustness.

### 2.1. Sentiment Analysis in the Tourism Domain

Sentiment analysis has become a valuable tool in the tourism industry, offering insights into traveler experiences through user-generated content [15, 16, 17]. By analyzing reviews and feedback, it helps stakeholders enhance services and understand public perception of destinations.

The researchers in [18] contributed a novel LDA topic-based sentiment analysis approach to analyze tourism reviews, combining topic modeling with lexicon-based sentiment analysis to extract insights from TripAdvisor reviews about Marrakech. In their methodology they used data scraping, pre-processing, LDA for the topic extraction, and sentiment analysis using the VADER and TextBlob, they achieved the accuracies of 77.3% and 72.6%, respectively, which outperformed the JST model by 3–7.7%. The study’s limitations include reliance on rule-based sentiment analysis, which struggles with irony and sarcasm, and a focus only on English reviews, limiting generalizability.

The SALSA project addresses the computational bottleneck in syntax-aware sentiment analysis by developing lightweight systems that combine fast syntactic parsing with explainable sentiment classification, enabling SMEs to perform accurate large-scale sentiment analysis without resource-intensive infrastructure [19]. The study introduces SEquence Labeling Syntactic Parser (SELSP), a method that treats dependency parsing as a sequence labeling task to accelerate syntax-based sentiment analysis while maintaining higher accuracy than conventional parsers (e.g., Stanza) and heuristic tools (e.g., VADER), making it viable for real-world SA applications[20].

In another research work, [21] contributed a novel deep learning-based approach to analyze sentiment and topics in tourism-related tweets during the Covid-19 pandemic, they focused on the hospitality and healthcare sectors. Their methodology combined VADER for sentiment analysis, LDA for topic modeling, and an LSTM-RNN model for sentiment classification, achieving the test accuracies of 80.9% for hospitality and 78.7% for healthcare. Their proposed model outperformed traditional machine learning methods like random forest and SVM, which demonstrated higher efficiency in capturing nuanced sentiments. However, the study had some limitations, such as reliance on Twitter data with API restrictions, which limited the dataset size, potentially affecting the LSTM’s performance. Additionally, the model struggled with sarcasm and irony, which are common challenges in sentiment analysis. Despite these shortcomings, the research provided a valuable insights into the public sentiment during the pandemic.

By combining BERT-based sentiment analysis with social media data from Twitter and Instagram to evaluate tourism in Granada, [22] contributed a novel approach. Their methodology involved the training of a Spanish-Tourism-BERT model for the sentiment classification, they used hashtags to identify the key tourist spots and their associated sentiments. The model achieved an accuracy of

75.7% for the Spanish texts, that outperformed the other classifiers, while Tweeteval was used for the English texts which gave the 75.6% accuracy. However, the study had limitations, such as data collection challenges due to Instagram's API restrictions and the need for larger training datasets to improve the sentiment analysis further. Despite these issues, the work provided valuable insights for tourism managers to enhance the destination marketing and services.

For Spanish tourism data, [23] introduced a novel multimodal sentiment analysis model, uniquely integrating both text and image inputs along with a data quality framework. Their methodology includes extracting opinions from social platforms, classifying sentiments separately in text and images using SenticNet 5 (adapted to Spanish) and facial recognition, and then combining the results through decision-level fusion. Their proposed model achieved 70% accuracy in text classification, 33% on the images, and 71% when both of the modalities were fused. However, the study acknowledges the limitations such as low image quality that impact facial emotion recognition and challenges with informal text from platforms like Twitter.

The researchers in [24] contributed a new context-aware, target-oriented method for sentiment and emotion analysis in tourism-related social media posts, focusing on specific aspects like attractions, accommodation, and food. They proposed a dictionary-based framework that combines manual annotation with lexicons, enabling more precise detection of emotions and sentiments linked to specific tourism targets. Although exact accuracy percentages are not provided, the system achieved high inter-annotator agreement (up to 92.3% for emotional words), showing strong reliability. However, the study is limited by the small annotated dataset of only 475 tweets, which may not fully capture the diversity of tourist opinions online.

Finally, [25] provided a comprehensive review of how sentiment analysis (SA) has been applied in tourism, offering a novel integration of bibliometric, systematic, and thematic analyses. They used VOSviewer to examine 111 papers from 2012 to 2021, clustering them into key research themes like consumer behavior, big data, and recommendation systems. While the study doesn't propose a specific model or report accuracy figures, it effectively maps the research landscape and identifies SA as a powerful tool in understanding tourist sentiment. However, the review they provided is limited by its reliance on the Scopus database and they have done it for English-language sources only.

## **2.2. Transformer-Based Approaches in Sentiment Analysis**

The introduction of transformer-based models has improved sentiment analysis by allowing a better understanding of context in text. Unlike traditional machine learning techniques that depend on manual features or basic word embeddings, transformer models like BERT, RoBERTa, and their multilingual versions learn bidirectional representations that capture subtle sentiment signals in different languages. These models perform well in various domains, including the tourism domain, where reviews often include informal and emotional language. Their ability to adapt through fine-tuning makes them suitable for the domain-specific tasks such as multilingual sentiment classification in tourism.

The study by Sudhir et al. [26] presented a clear comparison of sentiment analysis methods, including machine learning, deep learning, and transformer-based models. They tested several classifiers such as Naive Bayes, SVM, LSTM, and BERT on the IMDB dataset. Among these, the BERT model achieved the highest accuracy of 89.5%, while the BERT Large model with the UDA technique reached an accuracy of 95.22%. The paper also pointed out that the rule-based and lexicon-based approaches face challenges with complex text patterns such as sarcasm, and they often require frequent manual adjustments.

In another study, Bashiri et al. [27] performed a detailed comparison of transformer models used in sentiment analysis. They tested the BERT, RoBERTa, XLNet, ELECTRA, DistilBERT, ALBERT, T5, and GPT models on a total of 22 datasets. The authors followed a step-by-step approach to evaluate the accuracy and generalization ability of each model. Among all models, the T5 model gave the best results on most datasets, showing high flexibility and the ability to generalize well. The XLNet model was strong in detecting irony and product-related opinions, while the BERT and DistilBERT models gave the lowest performance, even though they were more efficient. The study also highlighted that the models still face problems with understanding sarcasm and idiomatic expressions, which are common

in user-generated reviews.

Two novel transformer-based models for explainable sentiment analysis were introduced in [28], focusing on generating extractive summaries to explain predictions. Their methodology involved a hierarchical transformer model (ExHiT) and a simpler sentence-based model (SCC), both applied to the IMDB dataset. The SCC model achieved the highest classification accuracy at 93.51%, while the ExHiT model reached up to 92.77% depending on the merging strategy. The study also explored explainability metrics, showing SCC performed best with 70.74% precision, but ExHiT showed improvement with enhancements like sentence masking. A noted limitation is that ExHiT sometimes extracted less interpretable summaries, especially without proper masking and embedding strategies.

In their research Wang et al. introduced a novel multimodal sentiment analysis model called TEDT, which uses a transformer-based encoder-decoder framework to fuse text, audio, and visual data [29]. The main innovation lies in converting nonnatural language features into natural language representations using a modality reinforcement cross-attention module and a dynamic filtering mechanism. Their model achieved high accuracy (89.3%) on the CMU-MOSI dataset and 85.9% on the CMU-MOSEI dataset, outperforming many existing methods. However, the TEDT model has high training time and struggles with accurately interpreting sarcasm and subtle expressions, which limits its performance in complex emotional scenarios.

A wide range of BERT-based transformer models is reviewed by [30] specifically for text-based emotion detection, providing insights into their performance, strengths, and weaknesses. They analyzed models like BERT, RoBERTa, XLNet, and DistilBERT across several datasets such as SemEval, EmotionLines, and ISEAR, highlighting how fine-tuned versions achieved strong results: for instance, HRLCE with BERT achieved an F1 score of 0.7709. The novelty lies in the comprehensive comparison of how different model architectures and training strategies affect emotion detection accuracy. However, the paper noted that models often struggle with detecting mixed or subtle emotions and face challenges like fixed input lengths and high computational costs.

A novel BERT-based CBRNN model for sentiment analysis on social media data, aiming to handle challenges like noisy text and contextual information loss, is proposed by [31]. The model combines zero-shot classification for data labeling, a pre-trained BERT for semantic embedding, dilated CNN for local and global feature extraction, and Bi-LSTM for capturing sequence dependencies. It achieved high accuracy, with 97% on the US-airline dataset and 93% on the IMDB dataset, outperforming other models in precision, recall, and AUC scores. However, the study noted the complexity of the hybrid model and the increased training cost as limitations.

In [32] AlBadani et al. introduced a novel Sentiment Transformer Graph Convolutional Network (ST-GCN) that models text data as a heterogeneous graph and learns sentiment-related node representations using transformer-based mechanisms. Their method combines BERT-based embeddings, graph structure with TF-IDF and PMI-based edges, and Laplacian eigenvector-based positional encoding. The ST-GCN model achieved high accuracy 95.43% on SST-B and 94.94% on IMDB outperforming several state-of-the-art models. However, the paper notes that the model's performance is slightly sensitive to the removal of low-frequency words and may require careful tuning of learning rates and epochs to avoid overfitting.

Overall, the existing body of work highlights the effectiveness of transformer-based models and data augmentation techniques in addressing sentiment analysis challenges, particularly in multilingual and domain-specific contexts. However, limited attention has been given to simultaneously handling class imbalance, multilingual variation, and multitask objectives within the tourism domain. This gap reinforces the need for more comprehensive and targeted approaches, such as the one proposed in this study, to improve generalization and performance across all subtasks in the Rest-Mex 2025 challenge.

### 3. Methodology

The methodology in this study consists of three sections; exploratory data analysis to under the distribution of the data instances, mitigating strategies to overcome the class imbalance issue and model

training and evaluation.

### 3.1. Exploratory Data Analysis

It's important to understand the structure and distribution of the individual instances in dataset as suggested in the study [33] that the quality of the dataset may impact the classification accuracy. We explored three key aspects: sentiment polarity, attraction type, and town-wise review instances. This analysis helped us identify potential issues like class imbalance, which could affect the performance of the model.

#### 3.1.1. Sentiment Polarity

The dataset is heavily skewed toward positive feedback. As seen in Figure 1, the "Very Positive (5)" label has the most entries with 136,561 reviews. The "Positive (4)" class follows with 45,034 instances. Neutral sentiments labeled as "3" are much fewer, totaling 15,519. Negative sentiments are rare—"Negative (2)" has 5,496 and "Very Negative (1)" has only 5,441 reviews.

This imbalance means the model might learn to favor positive predictions and ignore the less common negative or neutral ones. This is a common issue in sentiment datasets and needs to be handled carefully during pre-processing.

#### 3.1.2. Attraction Type

Figure 2 shows the number of reviews by attraction type. Restaurants are the most reviewed category with 86,720 instances. Next comes "Attractive" places like monuments or tourist sites with 69,921, and then hotels with 51,410 reviews. The distribution here is uneven but still fairly manageable compared to sentiment polarity. However, we still consider this variation when training the type classification model.

#### 3.1.3. Town-wise Distribution

Figure 3 highlights a clear imbalance in the number of reviews across different towns. Tulum has the highest count with 45,345 reviews, followed by Isla Mujeres (29,826) and San Cristóbal de las Casas (13,060). On the other end, towns like Tapalpa (725) and Real de Catorce (760) have very few samples.

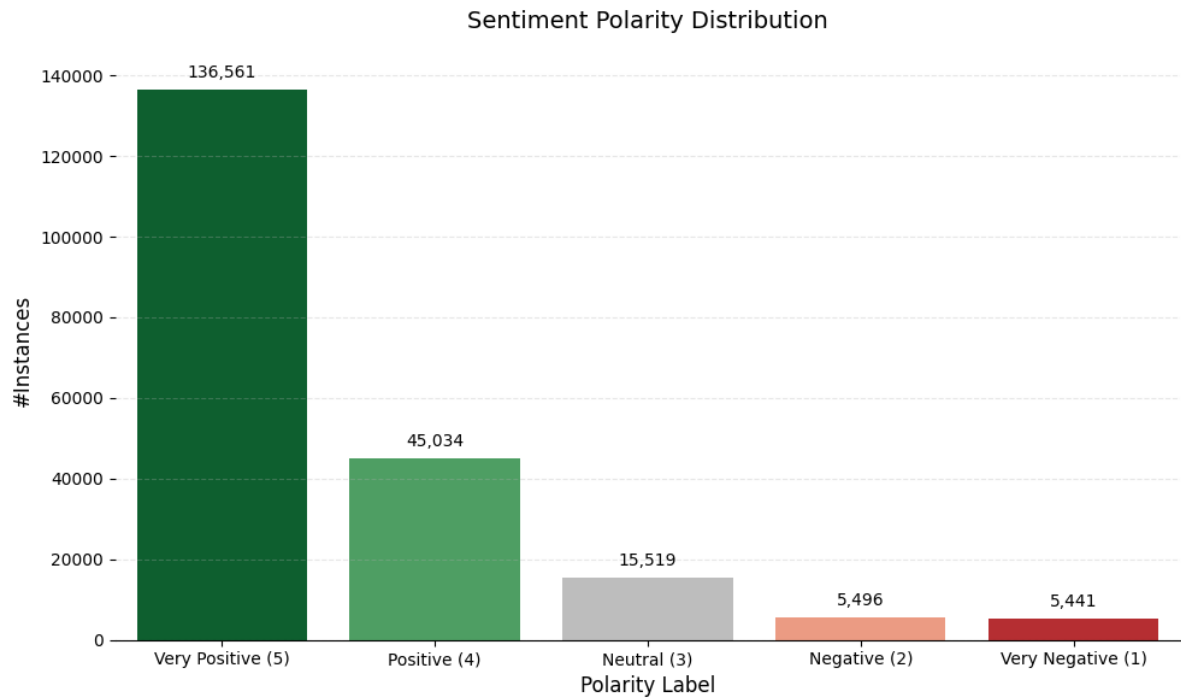
This makes town prediction quite challenging. Models usually need a balanced amount of training data for each class, and in this case, many towns are underrepresented. This imbalance can cause the model to perform poorly on towns with fewer examples, which is something we address later in the data augmentation phase.

To deal with this, we introduced specific data augmentation strategies to boost the presence of minority classes. These steps are explained in the next section, where we describe our approach to handle class imbalance issue for training more balanced models.

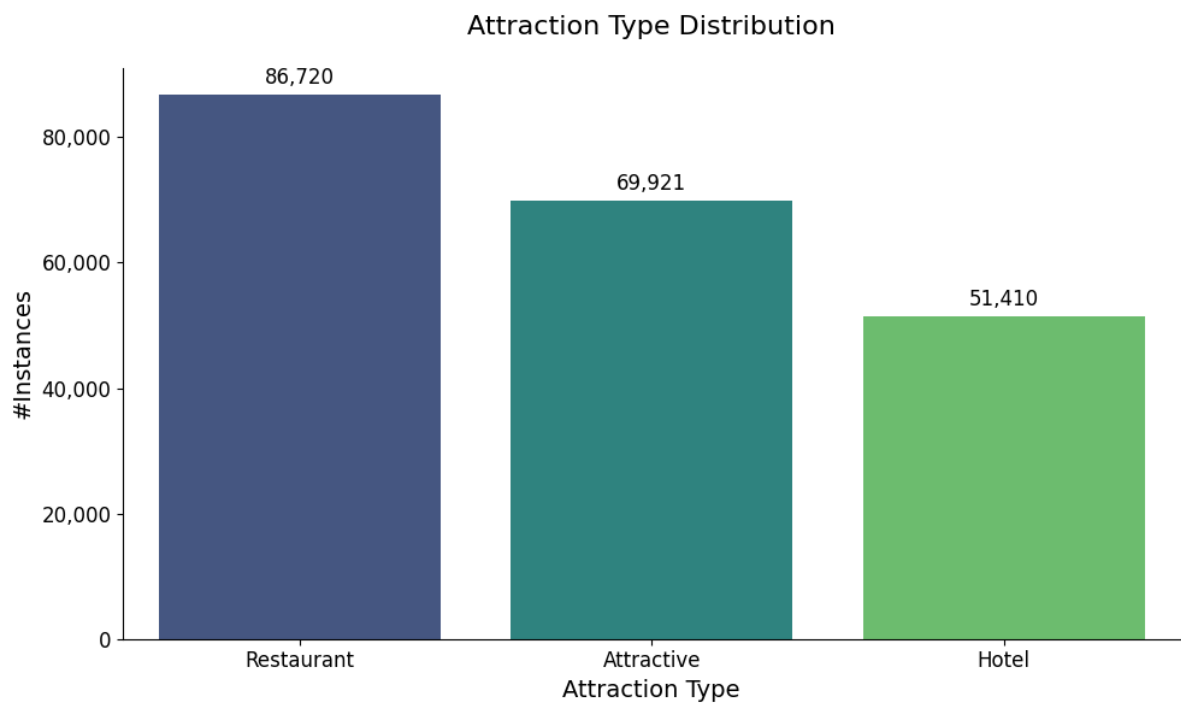
### 3.2. Mitigating Class Imbalance Issue

A key challenge in real-world datasets is class imbalance where unequal distribution across target classes hinders accurate model classification [34]. Standard Machine Learning (ML) approaches often fail on imbalanced data by optimizing majority class accuracy while neglecting minority classes [35]. We used the following three strategies to mitigate the class imbalance issue in the Rest-Mex 2025 training data split.

Figures 1, 2 and 3 present the distribution on instances based on sentiment polarity, attraction type and town in the training dataset respectively.



**Figure 1:** Sentiment polarity distribution in training dataset

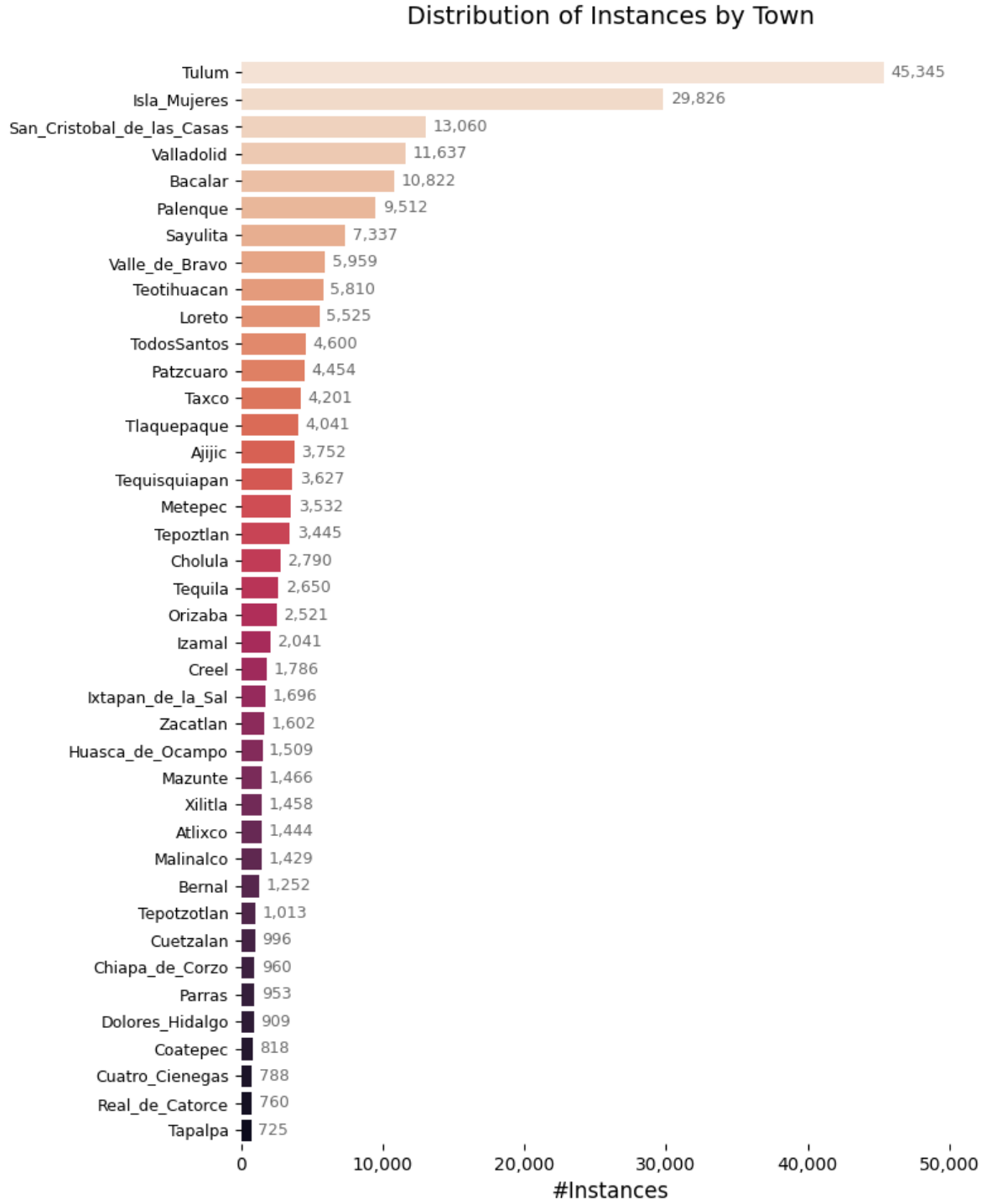


**Figure 2:** Distribution of attraction type in training dataset

### 3.2.1. Strategy 1: Oversampling

To mitigate class imbalance issue in the Rest-Mex 2025 training dataset, we used an oversampling technique to generate synthetic samples for minority classes by randomly duplicating existing instances. For each underrepresented class, our oversampling technique randomly selects instances (with replacement) and appends them to the dataset. Then, this oversampled dataset is shuffled to ensure randomness





**Figure 3:** Distribution of instances by town in training dataset

and prevent bias in model training. This approach enhances the representation of minority classes, improving classifier performance.

### 3.2.2. Strategy 2: Back-translation

We used the back-translation technique [36] to translate the reviews from the target language to a source language to enhance the representation of the minority class instances. There are some open source models available to perform automatic translation of the text. We translated the reviews of

**Table 1**  
Back-translation Models

Translation Model	Source Language	Target Language	Polarity Label
Helsinki-NLP/opus-mt-es-en	Spanish	English	1,2,3
Helsinki-NLP/opus-mt-en-es	English	Spanish	1,2,3
Helsinki-NLP/opus-mt-es-fr	Spanish	French	1,2
Helsinki-NLP/opus-mt-fr-es	French	Spanish	1,2
Helsinki-NLP/opus-mt-es-it	Spanish	Italian	1,2,3
Helsinki-NLP/opus-mt-it-es	Italian	Spanish	1,2,3
Helsinki-NLP/opus-mt-es-gl	Spanish	Galician	1,2
Helsinki-NLP/opus-mt-gl-es	Galician	Spanish	1,2

minority class instances in the training dataset to structurally similar target languages (Galician, Italian, French) and a structurally dissimilar target language (English), and then translated them back to its source language (Spanish). Three new instances are generated by each instance with label 1 using the English, French, and Italian languages as target languages. Using back-translation technique, we added four new instances for each instance having polarity label 1 or 2 and two new instance were added for each instance with polarity label 3. Table 1 presents the translation models used to perform back-translation of underrepresented classes in the training dataset along with source language and the target languages used by these models, and the polarity labels of the instances which were translated using these models.

### 3.2.3. Strategy 3: Hybrid (Oversampling + Back-translation)

This strategy combines the previous two strategies (Oversampling and Back-translation) to mitigate the class imbalance issue. More specifically, we apply the oversampling technique on the dataset created in strategy 2 using back-translation.

## 3.3. Model Training

Transformer models have revolutionized natural language processing (NLP) and sentiment analysis due to their ability to capture long-range dependencies and contextual relationships in text. In this model, the representation of each token is computed by attending to all other tokens in the sequence to capture long-range dependencies regardless of their distance. Unlike recurrent neural networks (RNNs) or convolutional neural networks (CNNs), transformer models leverage a self-attention mechanism to weigh the importance of different words in a sentence, which enables them to understand nuanced meanings more effectively [37, 38]. This architecture allows transformer models to process entire sequences in parallel, significantly improving computational efficiency and scalability on large datasets. Pretrained transformer models, such as BERT [39] and RoBERTa [40], further enhance performance by leveraging transfer learning enabling fine-tuning on specific tasks like sentiment analysis [39].

In sentiment analysis, transformer models excel in capturing subtle emotional cues, sarcasm, and context-dependent expressions. Models like RoBERTa, trained on vast corpora using masked language modeling, develop a deep understanding of language structure, making them particularly effective for sentence-level classification tasks [40]. Multilingual variants, such as XLM-RoBERTa, extend this capability across languages, enabling sentiment analysis in diverse linguistic contexts. Platforms like Huggingface [41] simplify access to these models, providing pretrained implementations that can be fine-tuned for domain-specific applications. The transformer models offer a robust and efficient solution for sentiment analysis outperforming traditional methods in accuracy and adaptability [39, 38] by combining contextual awareness with transfer learning.

For sentiment analysis, we fine-tuned two RoBERTa-based pretrained checkpoints



**Table 2**

Hyperparameters to fine-tune the transformer models

Hyperparameter	Value
Epochs	10
Batch Size	32
Learning Rate	2e-5
Weight Decay	0.01
Metric for best Model	f1
Evaluation Strategy	epoch

**Table 3**

Evaluation results of the sentiment polarity prediction models on validation data split

Model	Task	Dataset Strategy	Ev1. Loss	Ev1. F1 Score
twitter-xlm-roberta-base	Polarity	Strategy 1	0.6143	0.8208
	Polarity	Strategy 2	0.3472	0.9337
	Polarity	Strategy 3	0.1795	0.9694
roberta-base-bne	Polarity	Strategy 1	1.8016	0.8127
	Polarity	Strategy 2	0.3685	0.9153
	Polarity	Strategy 3	0.3264	0.9709
	Town	Strategy 1	0.1959	0.9860
	Type	Strategy 1	0.02598	0.9972

(roberta-base-bne<sup>1</sup> and twitter-xlm-roberta-base<sup>2</sup>) on Rest-Mex 2025 training datasets which were created using three strategies to overcome class imbalance issue as explained in section 3.2. The training dataset was split into a 90:10 ratio for training and validation of the models. Table 2 presents the hyperparameters used to train the models and Table 3 presents the evaluation results of the trained models on the validation data split and the Table 4 presents the results model evaluation results on test dataset.

**Table 4**

Evaluation results on test dataset

Model	Dataset Strategy	Sentiment Track Score	Polarity (F1)	Type (F1)	Town (F1)
twitter-xlm-roberta-base	Strategy 1	67.29	58.93	98.01	62.63
	Strategy 2	67.46	59.43	98.01	62.63
	Strategy 3	67.62	59.92	98.01	62.63
roberta-base-bne	Strategy 1	67.03	58.13	98.01	62.63
	Strategy 2	67.10	58.34	98.01	62.63
	Strategy 3	67.10	58.37	98.01	62.63

## 4. Conclusion

This paper presented a comprehensive framework to address the multifaceted Rest-Mex 2025 challenge, which involves sentiment polarity prediction, attraction type classification, and specific town identification from Spanish-language tourist reviews of Mexican Magical Towns. Recognizing the inherent

<sup>1</sup><https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne>

<sup>2</sup><https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base>

complexities such as linguistic nuances, informal expressions, and significant class imbalance in the dataset, we proposed and evaluated several data augmentation strategies.

Our core contribution lies in the successful application of a hybrid augmentation approach, combining oversampling with back-translation using both structurally similar and dissimilar languages. This hybrid strategy, when used to fine-tune transformer-based models, proved highly effective. The results demonstrate that this approach significantly mitigates the class imbalance issue, leading to improved performance and generalization across all three subtasks. Notably, the roberta-base-bne model fine-tuned with the hybrid dataset achieved an F1 score of 0.9709 for sentiment polarity prediction, and twitter-xlm-roberta-base also showed strong performance with an F1 score of 0.9694 using the same strategy.

The findings underscore the efficacy of targeted data augmentation in enhancing the robustness of advanced transformer models, particularly in specialized domains like tourism with unique dataset challenges. Future work could explore the integration of more diverse linguistic resources for back-translation, further data augmentation techniques based on paraphrasing, or investigate the adaptability of these models to other low-resource tourism contexts.

Our system achieved 4th place in the overall sentiment analysis track of the Rest-Mex 2025 shared task on sentiment analysis, competing against 35 participating teams. The system obtained an overall sentiment analysis score of 67.62, with the following detailed performance metrics; polarity prediction (F1 Macro): 59.92, type prediction (F1 Macro): 98.01 and town prediction (F1 Macro): 62.63. This competitive ranking demonstrates the robustness of our approach in sentiment classification across multiple subtasks.

## Acknowledgments

We acknowledge grants GAP (PID2022-139308OA-I00) funded by MICIU/AEI/10.13039/501100011033/ and ERDF, EU; LATCHING (PID2023-147129OB-C21) funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU; and TSI-100925-2023-1 funded by Ministry for Digital Transformation and Civil Service and “NextGenerationEU” PRTR; as well as funding by Xunta de Galicia (ED431C 2024/02), and CITIC, as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

## 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] H. Sun, Y.-W. Chen, L. Lin, Tensorformer: A tensor-based multimodal transformer for multimodal sentiment analysis and depression detection, *IEEE transactions on affective computing* 14 (2022) 2776–2786.
- [2] 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>.
- [3] R. Guerrero-Rodríguez, M. A. Álvarez Carmona, R. Aranda, A. P. López-Monroy, Studying online travel reviews related to tourist attractions using nlp methods: the

- case of guanajuato, mexico, *Current Issues in Tourism* 26 (2023) 289–304. URL: <https://doi.org/10.1080/13683500.2021.2007227>. doi:10.1080/13683500.2021.2007227. arXiv:<https://doi.org/10.1080/13683500.2021.2007227>.
- [4] J. Huang, J. Zhou, Z. Tang, J. Lin, C. Y.-C. Chen, Tmbl: Transformer-based multimodal binding learning model for multimodal sentiment analysis, *Knowledge-Based Systems* 285 (2024) 111346.
  - [5] 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>.
  - [6] 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).
  - [7] 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).
  - [8] 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.
  - [9] 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.
  - [10] M. A. Jahin, M. S. H. Shovon, M. Mridha, M. R. Islam, Y. Watanobe, A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets, *Scientific Reports* 14 (2024) 24882.
  - [11] Z. Yuan, W. Li, H. Xu, W. Yu, Transformer-based feature reconstruction network for robust multimodal sentiment analysis, in: *Proceedings of the 29th ACM international conference on multimedia*, 2021, pp. 4400–4407.
  - [12] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, A. Masnur, Optimizing sentiment analysis on imbalanced hotel review data using smote and ensemble machine learning techniques, *Journal of Applied Data Sciences* 6 (2025) 921–935.
  - [13] S. N. Almuayqil, M. Humayun, N. Jhanjhi, M. F. Almufareh, D. Javed, Framework for improved sentiment analysis via random minority oversampling for user tweet review classification, *Electronics* 11 (2022) 3058.
  - [14] A. Taheri, A. Zamanifar, A. Farhadi, Enhancing aspect-based sentiment analysis using data augmentation based on back-translation, *International Journal of Data Science and Analytics* (2024) 1–26.
  - [15] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancan 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.
  - [16] 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.
  - [17] 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>.
  - [18] T. Ali, B. Omar, K. Soulaïmane, Analyzing tourism reviews using an lda topic-based sentiment

analysis approach, *MethodsX* 9 (2022) 101894.

- [19] C. Gómez-Rodríguez, M. Imran, D. Vilares, E. Solera, O. Kellert, Dancing in the syntax forest: fast, accurate and explainable sentiment analysis with salsa, *arXiv preprint arXiv:2406.16071* (2024).
- [20] M. Imran, O. Kellert, C. Gómez-Rodríguez, A syntax-injected approach for faster and more accurate sentiment analysis, *arXiv preprint arXiv:2406.15163* (2024).
- [21] R. K. Mishra, S. Urolagin, J. A. A. Jothi, A. S. Neogi, N. Nawaz, Deep learning-based sentiment analysis and topic modeling on tourism during covid-19 pandemic, *Frontiers in Computer Science* 3 (2021) 775368.
- [22] M. S. Viñán-Ludeña, L. M. de Campos, Discovering a tourism destination with social media data: Bert-based sentiment analysis, *Journal of Hospitality and Tourism Technology* 13 (2022) 907–921.
- [23] J. Monsalve-Pulido, C. A. Parra, J. Aguilar, Multimodal model for the spanish sentiment analysis in a tourism domain, *Social Network Analysis and Mining* 14 (2024) 46.
- [24] A. Alaei, Y. Wang, V. Bui, B. Stantic, Target-oriented data annotation for emotion and sentiment analysis in tourism related social media data, *Future internet* 15 (2023) 150.
- [25] F. C. Manosso, T. C. Domareski Ruiz, Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review, *Journal of Tourism, Heritage & Services Marketing* 7 (2021) 16–27.
- [26] P. Sudhir, V. D. Suresh, Comparative study of various approaches, applications and classifiers for sentiment analysis, *Global Transitions Proceedings* 2 (2021) 205–211.
- [27] H. Bashiri, H. Naderi, Comprehensive review and comparative analysis of transformer models in sentiment analysis, *Knowledge and Information Systems* 66 (2024) 7305–7361.
- [28] L. Bacco, A. Cimino, F. Dell’Orletta, M. Merone, Explainable sentiment analysis: a hierarchical transformer-based extractive summarization approach, *Electronics* 10 (2021) 2195.
- [29] F. Wang, S. Tian, L. Yu, J. Liu, J. Wang, K. Li, Y. Wang, Tedt: transformer-based encoding–decoding translation network for multimodal sentiment analysis, *Cognitive Computation* 15 (2023) 289–303.
- [30] F. A. Acheampong, H. Nunoo-Mensah, W. Chen, Transformer models for text-based emotion detection: a review of bert-based approaches, *Artificial Intelligence Review* 54 (2021) 5789–5829.
- [31] S. T. Kokab, S. Asghar, S. Naz, Transformer-based deep learning models for the sentiment analysis of social media data, *Array* 14 (2022) 100157.
- [32] B. AlBadani, R. Shi, J. Dong, R. Al-Sabri, O. B. Moctard, Transformer-based graph convolutional network for sentiment analysis, *Applied Sciences* 12 (2022) 1316.
- [33] M. Imran, A. Ahmad, Enhancing data quality to mine credible patterns, *Journal of Information Science* 49 (2023) 544–564.
- [34] M. Altalhan, A. Algarni, M. T.-H. Alouane, Imbalanced data problem in machine learning: A review, *IEEE Access* (2025).
- [35] F. Thabtah, S. Hammoud, F. Kamalov, A. Gonsalves, Data imbalance in classification: Experimental evaluation, *Information Sciences* 513 (2020) 429–441.
- [36] M. K. M. Boussougou, P. Hamandawana, D.-J. Park, Enhancing voice phishing detection using multilingual back-translation and smote: An empirical study, *IEEE Access* (2025).
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [38] D. Jurafsky, J. H. Martin, *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (2025). URL: <https://web.stanford.edu/~jurafsky/slp3/>, online manuscript released January 12, 2025.
- [39] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.
- [40] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* (2019).
- [41] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: *Proceedings of the*

2020 conference on empirical methods in natural language processing: system demonstrations, 2020, pp. 38–45.

## **A. Online Resources**

The code is available in the following GitHub repository.

- [GitHub](#)