LKE-IIMAS team at Rest-Mex 2023: Sentiment Analysis on Mexican Tourism Reviews using Transformer-Based Domain Adaptation

Victor Giovanni Morales-Murillo[1,*], Helena Gómez-Adorno[2], David Pinto[1], Ilyan Alexey Cortés-Miranda[3] and Pierre Delice[1]

[1]*Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Language & Knowledge Engineering Lab (LKE), Puebla, México.*

[2]*Universidad Nacional Autónoma de México, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Ciudad de México, México.*

[3]*Universidad Nacional Autónoma de México, Facultad de Ciencias, Ciudad de México, México.*

Abstract

Sentiment analysis in tourist texts has gained relevance in the last decade because tourism is a social, cultural, and economic phenomenon that contributes to the economic development of countries. Thus, there is a need to create new natural language processing (NLP) mechanisms that improve tourism services and meet tourist needs. For this reason, LKE-IIMAS team participated in the sentiment analysis task of Rest-Mex 2023, which has three sub-tasks: identifying polarity, type, and country on a given opinion. For solving both sub-tasks, we first generated several training sets using stratified sampling on the training set of this shared task. Then, we applied a back-translation method as a data augmentation technique to tackle the data imbalance in the polarity sub-task. Furthermore, We used a masked language model based on RoBERTa that had been trained with a large Spanish dataset to generate three training strategies. Our first strategy was to fine-tune the base model for text classification. The second strategy was to adapt the base model to the tourism domain and to fine-tune the model adapted to text classification. The third strategy was to fine-tune the model adapted with the data augmented. Finally, our team LKE-IIMAS achieved first in the ranking on the sentiment analysis task at Rest-Mex 2023.

Keywords

Sentiment analysis, Mexican tourism, Transformers, Domain adaptation, NLP.

1. Introduction

This paper describes the participation of the LKE-IIMAS team in the sentiment analysis task at Rest-Mex 2023 [1]. The sentiment analysis in tourist texts has gained relevance in the last decade because tourism is a social, cultural, and economic phenomenon that contributes to the economic development of countries [2]. For example, this activity in Mexico represents 8.7% of the national gross domestic product (GDP), generating around 4.5 million direct jobs [3].

IberLEF 2023, September 2023, Jaén, Spain

*Corresponding author.

vg055@hotmail.com (V. G. Morales-Murillo); 0000-0002-6966-9912 (H. Gómez-Adorno); dpinto@cs.buap.mx (D. Pinto); alexcom@ciencias.unam.mx (I. A. Cortés-Miranda); padelice@gmail.com (P. Delice)

0000-0002-6786-9232 (V. G. Morales-Murillo); 0000-0001-7116-9338 (H. Gómez-Adorno); 0000-0002-8516-5925 (D. Pinto); 0009-0009-5363-6620 (I. A. Cortés-Miranda); 0000-0002-4170-4405 (P. Delice)

Thus, new mechanisms of natural language processing (NLP) are required to improve tourism services and meet tourist needs. This task is focused on the Spanish language because the most significant scientific works have focused on the English language [4], and some studies have focused only on peninsular Spanish. This task provides a text collection directly collected from tourists' opinions [5].

The sentiment analysis task of Rest-Mex 2023 consists in a classification task where the participating system is asked to predict the polarity of an opinion issued by a tourist who traveled to the most representative places, restaurants, and hotels in Mexico, Cuba, and Colombia. This collection was obtained from the tourists who shared their opinion on TripAdvisor between 2002 and 2022. Each opinion's class is an integer between [1, 5], where 1 represents the most negative polarity, and 5 is the most positive. Also, each opinion has a type label. The problem is defined as: "Given an opinion about a Mexican tourist place, the goal is to determine the polarity, between 1 and 5, of the text, the type of opinion (hotel, restaurant or attraction), and the country of the place of which the opinion is being given (Mexico, Cuba, Colombia)" [1].

Languages complexity hindered the classification of ironic, sarcastic, and subjective text on sentiment analysis [6]. Additionally, the models based on Long-short term memory (LSTM), approach fastText, convolutional neural networks (CNN), and transformers have been used as machine learning (ML) techniques in the literature [6] where the approach fastText has been faster than other ML models. The transformers model has achieved higher accuracy than other ML models. Besides, the best results in the previous editions of Rest-Mex were achieved by transformers approaches [3, 7, 8].

For this reason, we propose a transformer-based domain adaptation to improve sentiment classification accuracy in tourism reviews. This document is organized as follows. In section 2, we introduce the related works. In section 3, we describe a background on transformers. In section 4, we analyze the data. In section 5, we explain the proposed methodology in detail. In section 6, we show the results, and in section 7, we present our conclusions.

2. Related Work

Rest-Mex's previous editions encompassed a diverse range of methods for sentiment analysis classification. These methods included transformer-based models such as fine-tuned versions of Bert-like models, logistic regression classifiers, the Naive Bayes Multinomial algorithm, a cascade of binary classifiers, Bayesian techniques, KNN with Jaccard Distance, and an LDA-based approach for topic extraction. Additionally, a simple Deep Learning architecture was employed for classification.

In Rest-Mex 2021 [7], the team that achieved the best-performing approach employed two Bert-based strategies [9]. The initial strategy involved fine-tuning BETO, a pre-trained Bert-like model in Spanish. The second strategy combined Bert embeddings with TF-IDF weighted feature vectors. The second-best approach in 2021 involved utilizing the BETO model and a cascade of binary classifiers based on it [10]. The third-bet approach proposed an unsupervised keyword extraction technique to construct a sentiment weight list [11]. They employed SVM for classification, utilizing vector representations of text entities and matching the scores of prototypical words with the labels of the texts.

For the Rest-Mex 2022 evaluation campaign, the organizers used a new evaluation metric to assess that particular edition's sentiment analysis task. This metric is defined in equation 1.

$$\text{measure}_S = \frac{\frac{1}{1+\text{MAE}_p} + F_A}{2} \quad (1)$$

Where F_A is the average among the micro F-measure for each class, and MAE (see equation 2) evaluates the polarity.

$$\text{MAE}_{S_x} = \frac{1}{n}\sum_{i=1}^{n} |T(i) - S_x(i)| \quad (2)$$

Where S_x is a participating system x, $T(i)$ is the result of the instance i according to the Ground Truth, and $S_x(i)$ is the output of the participant system x, for example, i. Finally, n is the number of instances in the collection.

In Rest-Mex 2022 [3], the team UMU achieved the best results by approaching sentiment analysis as two distinct challenges: polarity classification as a regression problem and opinion target resolution as a multi-classification problem. Their pipeline included tasks such as text cleaning, extracting linguistic features, and utilizing non-contextual and contextual sentence embeddings using FastText, BERT, and RoBERTa models. They also performed fine-tuning and hyperparameter optimization, training neural network models with RMSE loss for regression and binary cross-entropy loss for classification [12]. The team that obtained second place, UC3M, explored two approaches. The first approach involved using SVM, a traditional machine learning technique, for text classification. The second approach utilized fine-tuned Transformers with different pre-trained models [10]. The third-place team, CIMAT, followed a methodology with two steps. They obtained high-level features from texts using an ensemble of BERT models trained for both subtasks. In the second step, they created an optimal ensemble of classifiers trained on the features obtained in the first step to make final predictions for each subtask. Their preprocessing involved minimal steps, including concatenating the title and opinion and converting the text to lowercase [11].

3. Background

Transformers became very popular in the Natural Language Processing (NLP) community, proposed by Vaswani et al. in 2017 [13], this neural architecture achieved and exceeded art state results for many NLP tasks. Its core mechanism is a self-attention process emphasizing the contextual relationships of words or tokens.

Transformers addressed the issue of long-term dependencies by utilizing a self-attention mechanism, which allows the model to weigh the importance of different words or tokens in the sequence while processing the entire sequence simultaneously. By self-attention mechanism, also known as scaled dot-product attention, the authors allow the model to compute the importance or attention weights for each word/token in the input sequence based on its relationships with other words/tokens. The attention allows the model to focus on relevant input parts during the encoding and decoding stages.

In this perspective, the use of self-attention mechanisms in Transformers demonstrates having the capability to perform well in situations of dependencies in the input space, both locally and on long-range dependencies. This capability has made Transformers highly effective in various NLP tasks, including sentiment analysis, where capturing contextual information and dependencies is crucial for accurate sentiment classification. However, some variants of Transformers are worth emphasizing in the context of our task: Bidirectional Transformers and Autoregressive Transformers, which are two different approaches within the Transformer architecture that serve different purposes.

The Bidirectional Transformers, such as BERT (Bidirectional Encoder Representations from Transformers), are designed to capture contextual information by considering both the left and right context of each word in the input sequence. They achieve this through masked language modeling, where some words in the input are randomly masked, and the model is trained to predict those masked words based on the surrounding context. BERT-based models have been widely used for various NLP tasks, including sentiment analysis, as they can effectively leverage the bidirectional context to understand the relationships between words [14].

On the other hand, autoregressive Transformers, such as GPT (Generative Pre-trained Transformer), focus on generating coherent and contextually relevant output by sequentially processing the input sequence, typically from left to right. The model predicts the next word in the sequence during training based on the preceding context. This sequential generation approach makes autoregressive models suitable for text generation and completion tasks. In sentiment analysis, autoregressive models can generate sentiment-related text, summarize sentiments, or generate responses based on given sentiments [15].

Contextual facts are an important piece of information for the sentiment analysis task. According to the literature [13] [16], the traditional approach based on bag-of-words or n-gram ignores the correlation structure between words by treating each word in an isolated manner. In other words, they ignore the dependency. However, the self-attention mechanism of transformer architecture considers the importance of context around each word. The self-attention mechanism allows this architecture to capture long-range dependencies and understand the sentiment expressed more nuancedly.

There is a wide variety of situations where transformers can improve sentiment analysis. They learn embeddings for each word in an unsupervised manner. These word embeddings encode semantic and syntactic information, enabling the model to capture more nuanced sentiment patterns and generalize better to unseen data.

For transfer learning, transformers can be pre-trained on large text corpora using tasks like masked language modeling or next-sentence prediction. This pre-training allows transformers to learn a general understanding of language, which can then be fine-tuned on specific sentiment analysis tasks with smaller labeled datasets. Transfer learning helps models leverage knowledge from diverse text sources, resulting in improved performance on sentiment analysis.

For this reason, the *RoBERTa (Robustly Optimized BERT Pretraining Approach)* [17] is used by our team due to this model is a variant of BERT. Furthermore, RoBERTa was trained on a much larger dataset with 160GB of text, which is more than 10 times larger than the dataset used to train BERT. On the other hand, RoBERTa has used a more effective training procedure because it utilizes a dynamic masking technique during training that helps the model learn more robust and generalizable representations of words.

4. Data Analysis

The Rest-Mex 2023 dataset contains 359,565 instances from tourists who shared their opinion on TripAdvisor between 2002 and 2022. This shared task dives its dataset into training and testing sets. The training set contains 251,702 instances which represent 70% of the original dataset, while the testing set contains 107,863 instances representing the remaining 30%. Each instance of the training set has a title, a review which is the opinion issued by the tourist, the polarity of the opinion (1, 2, 3, 4, 5), the country of the visited place (Mexico, Cuba, Colombia) and the type of place of which the opinion is being issued (Hotel, Restaurant, Attractive). At the same time, each instance of the testing set has an identifier, a title, and a review.

We analyze and process the training set of Rest-Mex 2023, which we divide into training and testing sets. The training dataset contains 226,531 instances that represent 90% of the training set of Rest-Mex 2023, while the test dataset contains 25,171 instances representing the remaining 10%. Figures 1, 2 and 3 present the three kinds of distribution of the training dataset: polarity, type, and country. It can be observed that this dataset is unbalanced, mainly observed in the polarity distribution. The majority class represented by label 5 of the polarity class has the 62.41% of the instances, whereas the minority class represented by label 1 only has the 2.29%. Further, the negative polarity values (1, 2) and the neutral polarity value (3) have a low number of instances in contrast to positive polarities (values 4 and 5). In the type distribution, the majority class is the Attractive label with the 44.17% of the instances, whereas the minority class is the Restaurant label with the 25.61%. In the country distribution, the majority class is the Mexico label with the 47.18% of the instances, whereas the minority class is the Cuba label with the 26.31%.

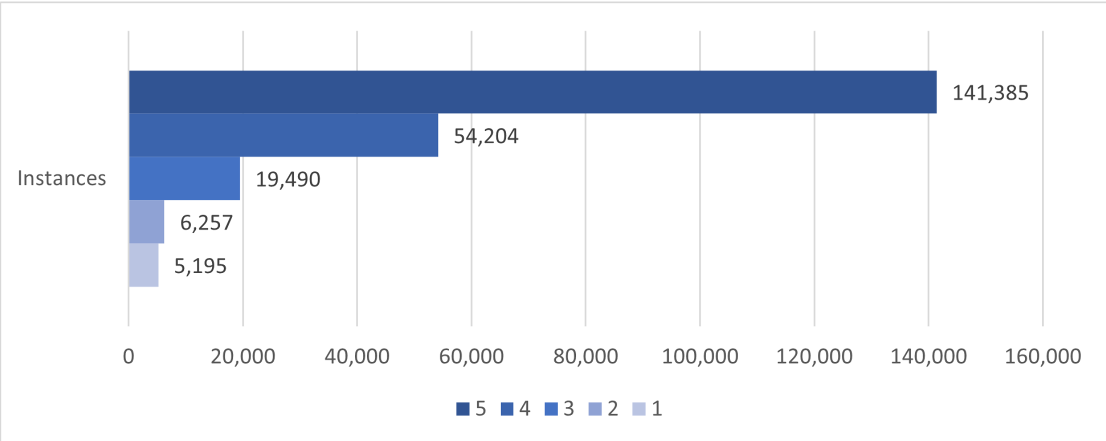

Figure 1: Distribution of the polarity class in training dataset.

5. Methodology

This section describes our methodology to tackle the sentiment analysis task at Rest-Mex 2023. Figure 4 presents the methodological design composed of 3 main steps: data augmentation,

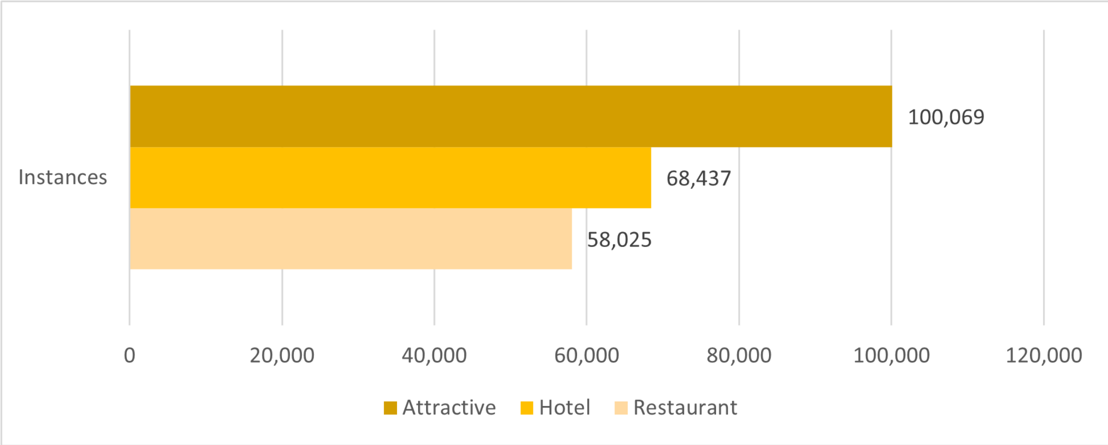

Figure 2: Distribution of the type class in training dataset.

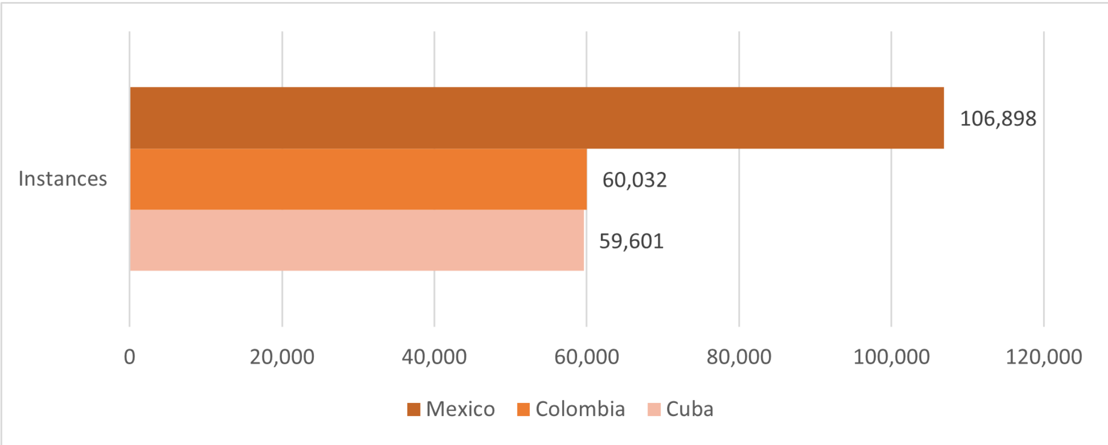

Figure 3: Distribution of the country class in training dataset.

domain adaptation, and training strategies. In the first step, we use the back-translation technique to tackle the data imbalance problem. In the second step, we perform a fine-tuning on a pre-trained language model of RoBERTa to adapt its data in a tourism domain. In the third step, we design three training strategies for a RoBERTa model based on the Spanish language. These steps are explained in more detail below.

5.1. Data Augmentation

Data imbalance in Machine Learning refers to an unequal distribution of dataset classes, and this problem is a challenge that needs more attention to resolve [18]. There are two main ways to tackle this problem: the undersampling and the oversampling methods. Undersampling techniques decrease the number of majority class instances, whereas oversampling methods increase the number of minority class instances creating new examples or repeating some.

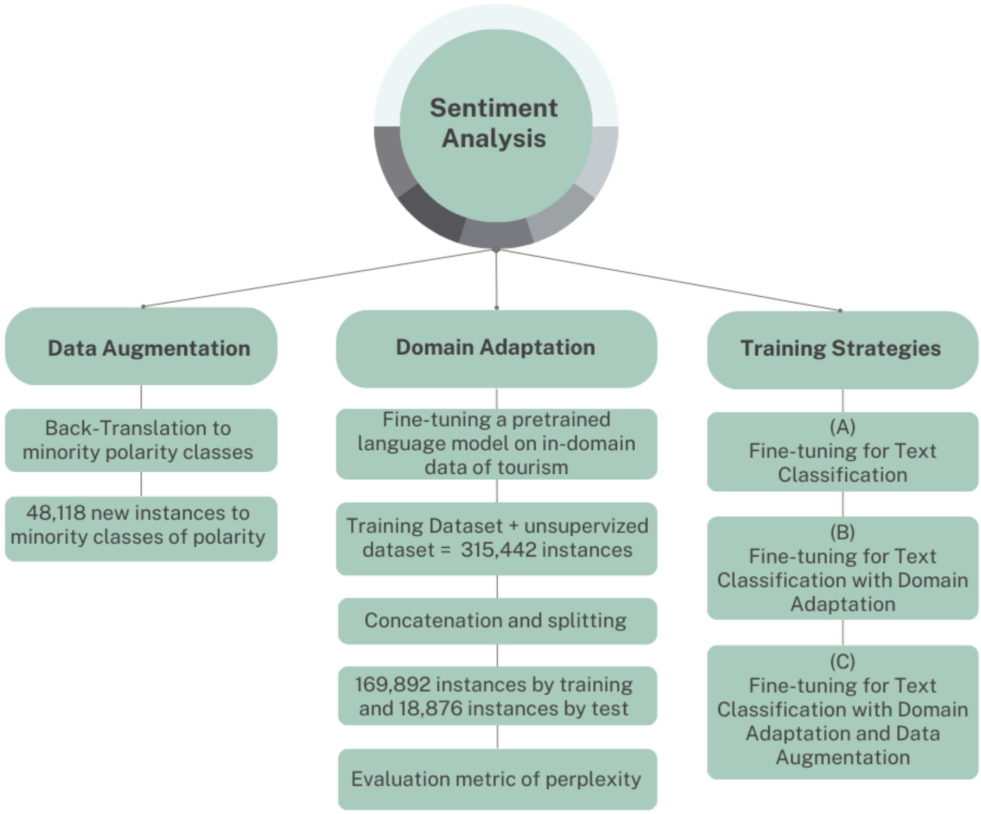

Figure 4: Methodological design.

Back-translation is an oversampling technique that translates a text from the target language to a source language [19]. There are several transformers open source models to perform automatic text translations.

We used an oversampling technique based on back-translation with transformers to generate new reviews by each instance with negative or neutral polarity values. We translate the text review of each instance of the training dataset to a target language, and the result is translated again to its source language, Spanish. So, several languages are employed as target languages. Three new instances are generated by each instance with label 1 using the English, French, and Deutsch languages as target languages. We generate two new instances by each instance with label 2 using the English and French languages as target languages. We produce one new instance by each instance with label 3 using as the target language the English language. Table 1 shows the transformers translation models used to perform back-translation of minority classes on the training dataset, the source language and the target language used by these models, and the labels of the instances translated by these models.

Figure 5 shows the new distribution of the polarity class in the training dataset, where 48,118 new instances are included and assigned to labels 1, 2, and 3. We generate a dataset with this new instances called analisis-sentimientos-textos-turisitcos-mx-polaridadV3-DA.

Table 1
Transformers translation models.

Model	Source language	Target language	Label
Helsinki-NLP/opus-mt-es-en	Spanish	English	1, 2, 3
facebook/nllb-200-distilled-600M	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-de	Spanish	Deutsch	1
Helsinki-NLP/opus-mt-de-es	Deutsch	Spanish	1

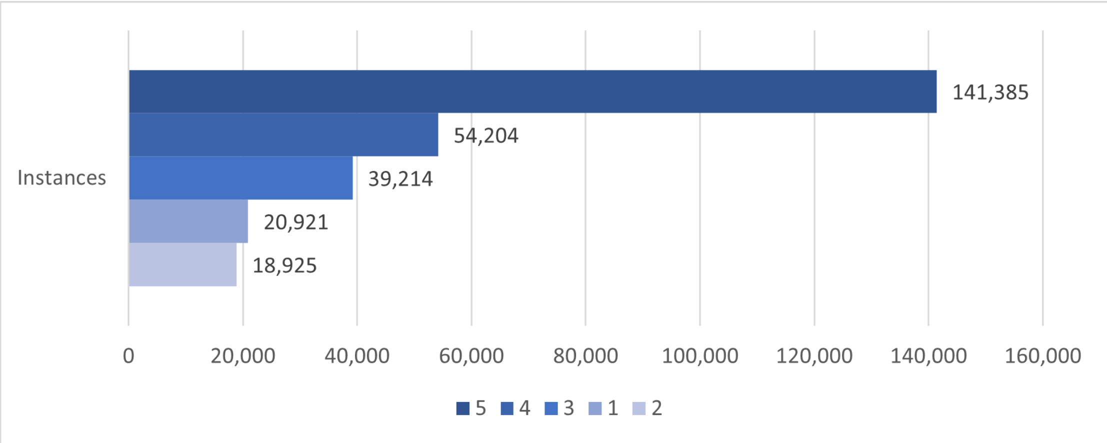

Figure 5: Distribution of polarity class in training dataset with data augmentation.

5.2. Domain Adaptation

Transformers models are used in many applications of NLP because they provide good results with transfer learning when the corpus used for pretraining a model is not too different from the corpus used for fine-tuning the specific task at hand. However, there are some cases when you want to fine-tune the language models on your data before training a task-specific head. So, this process of fine-tuning a pre-trained language model on in-domain data is called domain adaptation [16]. Such as, a legal contracts dataset hards the transfer learning with a transformer model as BERT because this model will process the domain-specific words of this dataset as rare tokens, and the performance results may not be satisfactory. For this reason, domain adaptation is an alternative for this problem, and it can increase the performance of many post tasks. It usually is achieved only one time [20].

We fine-tune a pre-trained model called *RoBERTa-base-bne* [21] that has been trained with a large corpus of Spanish with 570GB of clean and deduplicated text processed, which was gotten from a web crawling performed by the National Library of Spain from 2009 to 2019. We use an unsupervised dataset with 63,740 instances [9]. We join this unsupervised dataset with 251,702 instances of the training set of Rest-Mex 2023 to produce 315,442 instances of a new dataset that contains only text reviews without polarity, type, and country labels. In the preprocessing step, all instances' texts of the dataset are tokenized, concatenated, and divided into chunks of equal size to avoid that individual text might get truncated if their text is too long. Therefore, this step can reduce the losing information. The preprocessing result is a dataset with 188,768 tokenized instances with input_ids, attention_mask, word_ids, and labels as features. In fine-tuning, we randomly mask the 15% of the tokens in each batch of text as in the following example.

- <s>Me ha parecido un museo excepcional,<mask> en una zona<mask> accesible de la Ciudad de México. Las salas son impresionantes carpintería, sobre todo las <mask> Teotihu<mask>án y la de la cultura Maya<mask> Existe la posibilidad de que te acompañe un especialista<mask> la visita y<mask><mask> pena. <mask>…</s>;

As you can see, the <mask> token has been randomly inserted at various locations of text and our model should predict these tokens during training process. Furthermore, we divide the dataset preprocessed into training and testing sets, where the training datset contents 169,892 instances and the test dataset contens 18,876 instances. The main arguments for the training process are showed at Table 2.

Table 2
Training parameters for domain adaptation.

Parameter	Value
batch_size	64
evaluation_strategy	epoch
number_of_epochs	3
learning_rate	2e-5
weight_decay	0.01
per_device_train_batch_size	batch_size
per_device_eval_batch_size	batch_size
fp16	True
logging_steps	length of training dataset // batch_size

Perplexity [16] is a standard metric to evaluate the performance of language models due to that masked language models don't use datasets labeled, so this metric calculates the probabilities it assigns to the next word in all the sentences of the test dataset where high probabilities mean that model is not perplexed by the testing instances. We use the exponential of the cross-entropy loss as a mathematical definition of perplexity. We calculate the perplexity by evaluating the Transformers function, where a lower perplexity score means a better language model. The perplexity of our model without running the training is 28.03. In contrast, the perplexity of our model trained is 6.05, which is lower than the first result, representing good results for the domain adaptation on RoBERTa-base-bne. The masked language model result is called roberta-base-bne-finetuned-TripAdvisorDomainAdaptation.

5.3. Training strategies

Three main strategies for training are designed using a masked language model called RoBERTa-base-bne [21] that has been pre-trained using a large Spanish corpus with 570GB of clean and deduplicated text. Furthermore, the texts were obtained with a web crawling performed by the National Library of Spain from 2009 to 2019. We explain the strategies below.

In the first strategy, we fine-tune RoBERTa-base-bne for classifying polarity, type, and country on TripAdvisor review texts. We use of training set of Rest-Mex 2023 and divide it in a manner stratified by each class into training and test sets. The training dataset contains 70% of the training set instances of Rest-Mex 2023, whereas the test dataset contains the remaining 30%. Besides, we join the title and review of the instances as the only feature, rename the target class as labels, and remove the rest of the features. The result of this process is three datasets called analisis-sentimeinto-textos-turisitcos-mx-polaridad, analisis-sentimeinto-textos-turisitcos-mx-tipo and analisis-sentimeinto-textos-turisitcos-mx-pais (see https://huggingface.co/alexcom/). Then for the classification with Transformers, the polarity labels are modified by our team from [1, 2, 3, 4, 5] to [0, 1, 2, 3, 4], so the type labels are switched from [Attractive, Hotel, Restaurant] to [0, 1, 2] and the country labels are changed from [Mexico, Colombia, Cuba] to [0, 1, 2]. We tokenize each dataset with the RoBERTa-base-bne AutoTokenizer of Transformers to generate the features of labels, input_ids, and attention_mask. Moreover, we select the number of labels by each class. For example, the number of the labels to polarity classification is 5 and the number of the labels to type and country classification is 3. Besides, we use F1 metric for the model evaluation and show some training parameters for text classification in Table 3. Also, we use only 20,000 instances of the training dataset. We produce with this process three models called roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticos-mx-polaridad to polarity classification, roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticos-mx-tipo to type classification and roberta-base-bne-finetuned-analisis-sentimiento-textos-turisticos-mx-pais (see https://huggingface.co/vg055/) to country classification.

Table 3
Training parameters for text classification.

Parameter	Value
batch_size	16
num_train_epochs	2
evaluation_strategy	epoch
learning_rate	2e-5
weight_decay	0.01
per_device_train_batch_size	batch_size
per_device_eval_batch_size	batch_size
logging_steps	length of training dataset // (2*batch_size*num_train_epochs)

In the second strategy, we fine-tune roberta-base-bne-finetuned-TripAdvisorDomainAdaptation to classify polarity, type, and country on TripAdvisor review texts. We use of training set of Rest-Mex 2023 and divide it in a manner stratified by each class into training and test sets. The training dataset contains 90% of training set instances of Rest-Mex 2023, whereas the test dataset contains the remaining 10%. We process the dataset of same way that the first strategy and we generate three datasets called analisis-sentimientos-textos-turisitcos-mx-polaridadV2, analisis-sentimientos-textos-turisitcos-mx-tipoV2 and analisis-sentimientos-textos-turisitcos-mx-paisV2. Then, we perform the same steps from first strategy on the tokenizing, preprocessing and training but now with the model adapted to tourism domain and we now use all instances of training dataset for the training. We produce with this process three models called roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-polaridad to polarity classification, roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-tipo to type classification and roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-pais to county classification.

In the third strategy, we fine-tune roberta-base-bne-finetuned-TripAdvisorDomainAdaptation to classify polarity on TripAdvisor review texts. Unlike the other strategies, we use the dataset analisis-sentimientos-textos-turisitcos-mx-polaridadV3-DA, which is based on the training set of Rest-Mex 2023. Still, the training instances of negative and neutral classes were increased by a data augmentation technique applied previously. We repeat the process of the second strategy with this dataset and generate one model called roberta-base-bne-finetuned-TripAdvisorDomainAdaptation-finetuned-e2-RestMex2023-polaridadDA-V3 to polarity classification.

We performed in total 23 experiments, 19 experiments to polarity identification, 2 experiments to type classification and 2 experiments to country identification. We focused on polarity identification due to the problem is more complex for several reasons as the data imbalance and the classification of 5 classes. We included only the most representative results of these experiments by each training strategy on the next section.

6. Results

The results obtained during the development period are presented in Figure 6 where the second strategy presents the best result, followed by the third and first strategies. Polarity, type, and country classification results related to the first strategy are improved with the domain adaptation applied to the second strategy. However, the data augmentation on the third strategy does not represent an improvement to the second strategy, whereby it is essential to review the data augmentation techniques in more detail for future work.

We generate three outputs based on these strategies and the test set of Rest-Mex 2023. All outputs use the second strategy for the classification of type and country. Output 1 utilizes the first strategy for polarity classification, output 2 employs the second strategy for polarity classification, and output 3 applies the third strategy to classify the polarity. We send these outputs to Rest-Mex 2023, and the results gotten on the test process of Rest-Mex 2023 are presented in Figure 7, where the best result is achieved by output 2, followed by output 1 and output 3. However, the difference between the three outputs is very low due to only the polarity classification differentiates these outputs. On the other hand, we observe that domain adaptation is an excellent alternative to improve our results. However, looking for other methods to tackle the unbalance problem of the polarity classification is essential.

We achieved first place in the ranking with output 2 on the sentiment analysis task of Rest-Mex 2023, where 18 teams from around the world participated. Besides, our three outputs were ranked top three in this shared task.

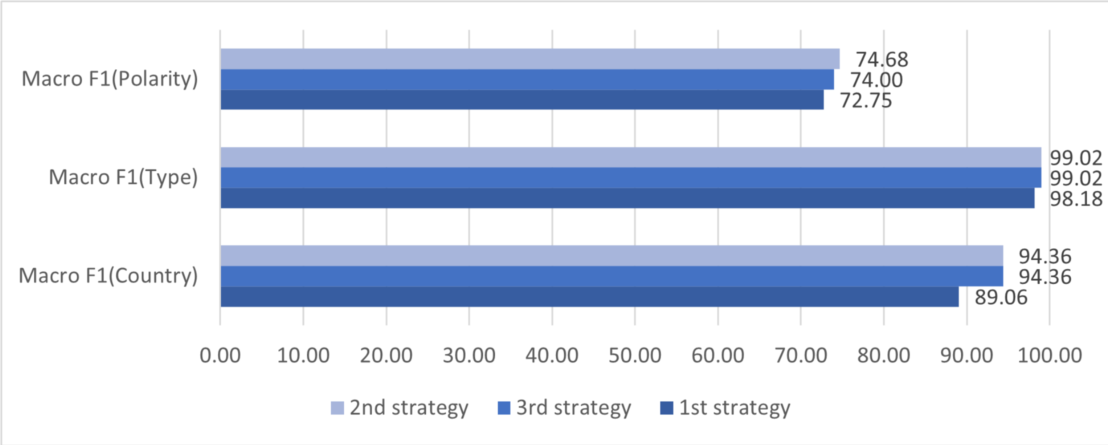

Figure 6: Results of training strategies

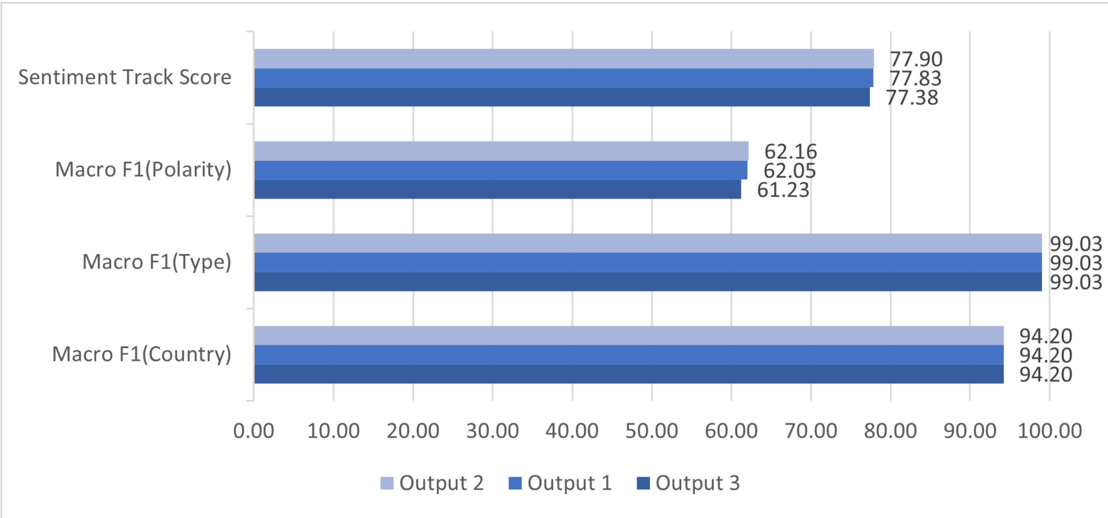

Figure 7: Results of outputs submission to Rest-Mex 2023

7. Conclusions

Our team, called LKE-IIMAS, achieve the highest results on the sentiment analysis task at Rest-Mex 2023 with a domain adaptation to a pre-trained model based on RoBERTa trained with a large dataset in the Spanish language. We developed three main steps in our methodology: data augmentation, domain adaptation, and training strategies. Besides, we generated several datasets with different types of stratification by classifying polarity, type, and country. We fine-tuned a masked language model to apply the domain adaptation and evaluated their perplexity. We fine-tuned a language model for polarity, type, and country text classification. We evaluated the result of these models with the F1 metric to select the three best strategies to generate three output submissions to Rest-Mex 2023. Domain adaptation improved the results. However, it is important to analyze other data augmentation methods that produce new instances with quality.

Acknowledgments

This work has been carried out with the support of DGAPA-UNAM PAPIIT project number TA101722. The authors also thank the Language & Knowledge Engineering Lab (LKE) of Benemérita Universidad Autónoma de Puebla, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) of Universidad Nacional Autónoma de México and CONACYT for the computing resources provided through the Deep Learning Platform for Language Technologies of the INAOE Supercomputing Laboratory. We also want to thank Eng. Roman Osorio for supporting the student administration of the project.

References

[1] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, A. Ramón, 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).

[2] 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. URL: https://doi.org/10.1080/0952813X.2022.2153276. doi:10.1080/0952813X.2022.2153276. arXiv:https://doi.org/10.1080/0952813X.2022.2153276.

[3] M. A. Álvarez Carmona, Ángel Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. G.-R. y Lázaro 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) 289–299. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6449.

[4] 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.

[5] CIMAT, Rest-Mex: Research on Sentiment Analysis Task for Mexican Tourist Texts, 2023. URL: https://sites.google.com/cimat.mx/rest-mex2023/.

[6] S. Balcí, G. M. Demirci, H. Demirhan, S. Sarp, Sentiment Analysis Using State of the Art Machine Learning Techniques, in: C. Biele, J. Kacprzyk, W. Kopeć, J. W. Owsiński, A. Romanowski, M. Sikorski (Eds.), Proceedings of MIDI'2021 – 9th Machine Intelligence

and Digital Interaction Conference, Springer International Publishing, Warsaw, 2022, pp. 34–42.

[7] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cardenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism, Procesamiento del Lenguaje Natural (2021).

[8] 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.

[9] J. Vásquez, H. Gómez-Adorno, G. Bel-Enguix, Bert-based approach for sentiment analysis of spanish reviews from tripadvisor, CEUR WS Proceedings, 2021, pp. 163–172.

[10] A. Ballester-Espinosa, Cascade of biased two-class classifiers for multi-class sentiment analysis., in: IberLEF@SEPLN, CEUR-WS.org, 2021.

[11] Geovanni Velazquez Medina, Delia Irazú Hernández Farías, Dci-ug participation at rest-mex 2021: A transfer learning approach for sentiment analysis in spanish., in: IberLEF@SEPLN, CEUR-WS.org, 2021.

[12] J. A. García-Díaz, M. Á. Rodríguez-García, F. García-Sánchez, R. Valencia-García, Umuteam at REST-MEX 2022: Polarity prediction using knowledge integration of linguistic features and sentence embeddings based on transformers, in: M. Montes-y-Gómez, J. Gonzalo, F. Rangel, M. Casavantes, M. Á. Á. Carmona, G. Bel-Enguix, H. J. Escalante, L. A. de Freitas, A. Miranda-Escalada, F. J. Rodríguez-Sanchez, A. Rosá, M. A. S. Cabezudo, M. Taulé, R. Valencia-García (Eds.), Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), A Coruña, Spain, September 20, 2022, volume 3202 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3202/restmex-paper13.pdf.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, ArXiv abs/1810.04805 (2019).

[15] A. Katharopoulos, A. Vyas, N. Pappas, F. Fleuret, Transformers are rnns: Fast autoregressive transformers with linear attention, in: International Conference on Machine Learning, 2020.

[16] L. Tunstall, L. von Werra, T. Wolf, Natural Language Processing with Transformers, Revised Edition, O'Reilly Media, 2022. URL: https://books.google.com.mx/books?id=8plxEAAAQBAJ.

[17] 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, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.

[18] R. Mohammed, J. Rawashdeh, M. Abdullah, Machine learning with oversampling and undersampling techniques: Overview study and experimental results, in: 2020 11th International Conference on Information and Communication Systems (ICICS), 2020, pp.

243–248. doi:10.1109/ICICS49469.2020.239556.
[19] N. L. Pham, V. Vinh Nguyen, T. V. Pham, A data augmentation method for english-vietnamese neural machine translation, IEEE Access 11 (2023) 28034–28044. doi:10.1109/ACCESS.2023.3252898.
[20] Y. Zhu, Y. Qiu, Q. Wu, F. L. Wang, Y. Rao, Topic driven adaptive network for cross-domain sentiment classification, INFORMATION PROCESSING & MANAGEMENT 60 (2023). doi:10.1016/j.ipm.2022.103230.
[21] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, A. Gonzalez-Agirre, C. Armentano-Oller, C. Rodriguez-Penagos, M. Villegas, Spanish language models, 2021. arXiv:2107.07253.