Transformers Approach for Sentiment Analysis: Classification of Mexican Tourists Reviews from TripAdvisor

Mario Pérez Enríquez[1], Javier Alonso Mencía[1] and Isabel Segura-Bedmar[1]

[1]*University Carlos III of Madrid, Spain*

Abstract

Sentiment analysis is one of the coveted areas of Natural Language Processing due to its business application with a scope that covers any type of opinion based text. In this paper, we describe our participation at the track of sentiment analysis of the REST-MEX@IberLef 2022 shared task. The objective of this sub-task is to predict the polarity of tourists' opinions about representative places in Mexico. Most opinions are written in Spanish, though there also some opinions written in English. Nevertheless, the number of opinions written in English is not meaningful. We tackle the task using traditional machine learning classifiers as well as fine-tuned Transformers for sentiment analysis. Throughout our experimentation, we also apply novel data augmentation techniques (such as summarization) to improve the result of the models. Our experiments show that the fine-tuned transformer models can obtain successful results ranking the second place at the REST-MEX@IberLef 2022 shared task.

Keywords

Sentiment analysis, REST-MEX, Transformers, NLP.

1. Introduction

Sentiment analysis aims to study on how opinions are expressed and identify their polarity (positive, negative or neutral) towards a specific topic. [1] Many organizations and companies can benefit from the use of automatic sentiment analysis technology, since it allows them to efficiently mine the opinions about a specific subject in a large number of comments. Tourism is likely one of the sectors that can obtain a greater benefit from sentiment analysis. This is mainly due to tourists are especially prolific in sharing their opinions on social networks. Thus, social media is having an tremendous impact on tourism [2].

Mexico is one of the countries worldwide where tourism generates the greatest impact on the generation of direct employment in that sector, and therefore, economic growth to achieve stability among its inhabitants [3]. In 2018, more than 4.5 million direct jobs were generated in Mexico thanks to tourism, corresponding to 8.5% of the GDP of this country [4]. In the last two years, Covid-19 has negatively influenced global tourism, leading to very negative consequences for developing countries' economies.

To overcome economic crisis after Covid-19, it is vital to increase productivity on sectors such as tourism [5]. Thanks to tourism platforms such as Tripadvisor [6], where people can post their opinions about places they visit, millions of data are generated in text form that could

IberLEF 2022, September 2022, A Coruña, Spain

be analyzed for various purposes. For example, to understand why some areas tend to receive more tourists than others. The rapid growth of available data together with the development of new efficient techniques for Natural Language Processing (NLP) [7], makes it possible to perform analysis of large volumes of data.

Sentiment analysis is a task within the field of NLP that examines subjectivity. Distinguishing subjective texts is a complex problem, even for individuals, as it is a combination of emotions, opinions, speculations or irony among others [8].

In recent years, several NLP based approaches have been exposed to solve the sentiment analysis task applied to different topics [9, 10, 11]. Early approaches [12], [13] for sentiment analysis exploited classical machine learning algorithms such as SVM or Naïve Bayes and text representation models like Bag-of-Words, Term Frequency-Inverse Document Frequency (TF-IDF). Due to the recent advances in computing power [14], traditional machine learning models are being replaced by newer techniques based on deep neural networks, which are obtaining state-of-the-art results in many NLP applications. [15, 16]. The big evolution in the NLP field comes from 2017 with the introduction of Transformers [17], a novel technique based on neural networks that offers superior results to previous ones due to its ability to understand context to a greater extent. Since Transformers enable data paralelization, they are usually more efficient than previous deep learning models such as recurrent neural networks [17].

In this paper, we propose the use of Transformers methods to solve the sentiment analysis sub-task of the REST-MEX@IberLef 2022 shared task [18], preceding the 2021 version [19]. This subtask aims to identify the level of satisfaction as well as the type of place that a tourist's opinion is describing about a Mexican place. The polarity of an opinion is defined in the numerical interval 1-5 (1 being the worst and 5 the best). The type of place can be a "Restaurant", "Hotel" or "Tourist attraction".

2. Methodology

2.1. Data analysis

The given training dataset consists of a total of 30,212 comments. Each comment contains the following fields: title, opinion, attraction (type of place) and polarity.

First, we need to identify if the training dataset is well-balanced, that is, all classes have similar number of instances. An unbalanced dataset means that the models will not be able to train properly on the least represented classes, which will hindered their inference as there is not enough insight to classify them.

The analysis of polarity distribution on Figure 1 shows that polarity is highly biased towards the most positive polarity (value 5), representing 70% of the polarity values, meanwhile the lowest polarity (value 1) has a representation of less than 2%.

In terms of the type of place (attraction), it is more evenly distributed with "Hotel" holding 55% of the values and the other half divided between the rest.

We also need to analyze the content of the text fields: title and opinion. Most texts contain emojis specially prevalent in the titles. As shown in table 2, some titles include only a few tokens (in the form of emojis) with an average of four tokens. In terms of the opinions, most of the instances have around 106 tokens which is fair but some are as long as 3,484.

Polarity values (30212)		Attraction values (30212)	
Class	Instances	Class	Instances
5	20,936	Hotel	16,565
4	5,878	Restaurant	8,450
3	2,121	Attractive	5,197
2	730	-	-
1	547	-	-

Table 1
Initial analysis of instances for Polarity and Attraction classes

Opinion field - Number of tokens		Title field - Number of tokens	
Count	30,210	Count	30,210
Mean	106.24	Mean	3.99
Std.	162.73	Std.	2.78
Min.	3.0	Min.	1.0
25%	35.0	25%	2.0
50%	53.0	50%	3.0
75%	104.75	75%	5.0
Max.	3484	Max.	26.0

Table 2
Analysis of tokens in Opinion and Title fields

Preprocessing is performed to remove opinions with more than 5,000 characters, 278 long opinions were removed due to performance reasons while training models. On the other hand, 2 instances were removed for containing empty fields.

2.2. Data Augmentation

Due to the unbalanced nature of the dataset, data augmentation was performed to increase the instances of the more underrepresented classes.

Unlike data sampling it consists of duplicating o removing instances from the dataset, data augmentation stands to create new instances by applying different type of techniques to the existing dataset.

Data augmentation for NLP is a complicated task specially for text classification as you do not want to change the meaning of the sentence and also you cannot duplicate records. Otherwise you run the risk of overfitting some labels to specific data characteristics [20].

Data augmentation was only applied to increase the number of instances in the unrepresented polarity classes.

Various techniques exists to solve this issue and the one picked is summarization. Summarization is the task of generating shorter input text while preserving the important information [21]. The idea behind using summarization for data augmentation is that the meaning of the sentence is not changed, so that text meaning characteristics remain the same without it being a duplication. The summarization model used [22] is a pretrained model based of the mT5 transformer architecture [23].

Text summarization models can propose different summaries from the same input depending on the desired length of the text output. Taking this fact into account, for polarity 1 we generate 3 different summaries, for polarity 2 we generate 2 different summaries and for polarity 3 we generate 1 different summary. In this way, we achieve to increase the number of instances in the three most unrepresented classes (polarities 1, 2 and 3).

Once the data augmentation is performed, the resulting training dataset consists of 38,551 instances (that is, 8341 were automatically generated), as shown in Figure 3 (polarity distribution). The polarity 5 drops to 50% of the representation while polarity 1 increases to 7% compared to original.

Polarity values (38832)		Atttraction values (38832)	
Class	**Instances**	**Class**	**Instances**
5	20936	Hotel	20058
4	6363	Restaurant	11450
3	5878	Attractive	7324
2	2920	-	-
1	2735	-	-

Table 3
Analysis of instances for Polarity and Attraction classes after data augmentation

2.3. Scenarios for training

As the training dataset contains two text fields, we proposed different combinations of these fields to create the input texts for training:

1. For each instance, we only exploit the opinion field and discard the title one. From now, we call it as "Dataset 1".
2. We apply data augmentation on the Dataset 1. From now, we called it as "Dataset 2".
3. For each instance, the title and opinion fields are concatenated to form the input text. From now, we call it as "Dataset 3".
4. We apply data augmentation on the Dataset 3. From now, we call it as "Dataset 4".

The reason to discard the title field in some scenarios is that this field might not give extra information considering the high prevalence of emojis in the titles, which our experimentation models will handle as unknown tokens.

In each dataset, 80% of data is used for training and the remaining 20% for validation. Table 4 shows some statistics about each of scenarios created for training:

2.4. Approaches

In order to solve the given task, two different approaches have been followed. The first approach is a traditional machine learning technique known as SVM (Support vector machine) [24], which has been successfully applied for text classification tasks.[25] Our second approach is based on fine-tuned Transformers, where we explore different pretrained models.

Name	Fields	Data augmentation	Train instances	Validation instances
Dataset 1	Opinion	No	23,944	5,987
Dataset 2	Opinion	Yes	30,840	7,711
Dataset 3	Title + Opinion	No	23,944	5,987
Dataset 4	Title + Opinion	Yes	30,840	7,711

Table 4
Description of the four scenarios for training

2.4.1. SVM

is a supervised learning algorithm whose objective is to find a hyperplane that best separates different classes in the form of data points. We consider this classifier as our baseline.

For this method, it was necessary to convert each input text to a embedding, that is, a numerical vectorized representation. In order to do that, a Spanish word embedding model, containing almost 3 billions tokens, was used with a vector dimension of 300 [26]. Its implementation is based on FastText and SUC (Spanish Unannotated Corpora) vocabulary [27].

To train our SVM baseline, Python library `sklearn` provides an easy way to implement, train an test an SVM classification model (SVC) in few lines of code.

2.4.2. Transformers.

Over the last few years, there have been tremendous advances in applying Transformers to many NLP applications. This is mainly due to the existence of platforms such as Huggingface [28], which provides a simple way to use and train pretrained models. Using a pretrained model means we can use transfer learning so that the model is not trained from scratch, saving time, resources and yielding better results [29]. Huggingface models make this task simple, as it provides implementations for the different NLP tasks.

The following list contains the pretrained models that we fine-tuned for the proposed task:

- **RoBERTa**: The pretrained model selected is a specific version of RoBERTa [30] trained with Spanish vocabulary called RoBERTaESP [31]. The text is tokenized with the same tokenizer from the pretrained model. RoBERTa is a highly optimized BERT[16] model, sharing a very similar architecture. Moreover, RoBERTa has been trained using a masked language modeling objective, that is 15% of the tokens are masked. It implements the encoder part of the transformer architecture [17] and it is aimed for tasks that use whole sentences to make decisions.
- **GPT2**: The pretrained model selected is a specific version of GPT2 [32] trained with Spanish vocabulary called `GPT2 base bne` [31]. The text is tokenized with the same tokenizer from the pretrained model. GPT2 it has been trained with a casual language modeling objective, that is each future token is masked. It implements the decoder part of the transformer architecture and sit is an auto regressive model with a focus on text generation.

With the pretrained models RoBERTa and GP2T we apply a text classification head using the Huggingface API (`AutoTokenizer, TFAutoModelFor SequenceClassification,` DataCollatorWithPadding classes) to fine-tune the models with our own datasets.

We picked these two models as they represent the two parts of the Transformers architecture in encoder and decoder. Encoder models perform better at text classification and token classification techniques while decoder models outperform encoders at text generation [33].

For the training these models, we used AdamWeigthDecay with a learning rate of 2e-5, decay rate of 0.01, no warm up steps, 2 epochs and number of steps equal to the number of items on the dataset multiplied by the epochs. The rest of the parameters follow the defaults as in Huggingface's implementation.

3. Results and Discussion

3.1. Models comparison

This subsection shows the results obtained for each proposed approach: SVM and RoBERTaESP are evaluated in detail on the four scenarios described above. However, the GPT2-base-bne) approach is only evaluated on Dataset 1, because no improvement in results was gained from RoBERTaESP on Dataset 1.

Table 5 shows the results for the task of identifying the type of place (attraction). All models tested for this task produced high accuracy results. However, the RoBERTaESP slightly improves the result with respect to the rest models, reaching 98% accuracy. Data augmentation does not have an effect because the focus was for the underrepresented polarity classes. It can be observed that SVM performs worse on datasets with title and opinion while RoBERTa remains the same. Therefore RoBERTa outperforms the other models for this attraction prediction.

Attraction classification on validation dataset					
Model	**Dataset**	**F1 - Hotel class**	**F1 - Restaurant class**	**F1 - Attractive class**	**Accuracy**
SVM	1	0.96	0.92	0.97	0.95
RoBERTaESP	**1**	**0.99**	**0.98**	**0.99**	**0.98**
GPT2	1	0.98	0.97	0.99	0.98
SVM	2	0.95	0.91	0.95	0.94
RoBERTaESP	**2**	**0.99**	**0.98**	**0.99**	**0.98**
SVM	3	0.97	0.93	0.97	0.96
RoBERTaESP	**3**	**0.99**	**0.98**	**0.99**	**0.98**
SVM	4	0.96	0.92	0.96	0.95
RoBERTaESP	**4**	**0.99**	**0.98**	**0.99**	**0.98**

Table 5
Results on Attraction task on validation dataset

Table 6 shows the results from the task of identifying the polarity on the validation dataset. This task shows more varied results among the different models and scenarios. First of all, it is important to recall that this task has five different classes which represent the polarity of a review, from 1 (most negative polarity) to 5 (most positive polarity). Sometimes it can be difficult, even to humans, to differentiate when a bad opinion should get 1 or 2 stars (for example, "This hotel was very bad"). That is, some similar opinions could be labeled slightly different.

Regarding the different tasks, SVM shows a similar performance comparing with Transformers for the Attraction task. However, when the number of classes increases (Polarity task), there are significant changes. First, SVM is not able to classify almost any instance into intermediate classes (polarities 2 and 3). Secondly, although Transformers and SVM reach similar accuracy values, F1-values state a more balance classification from Transformers models. Thirdly, among the Transformers models, RoBERTaESP gets the best overall accuracies values and most of the F1-values in both tasks (Attraction and Polarity). Finally, Classes 2 and 4 from the Polarity task have been the most difficult ones to classify, even for the best RoBERTaESP model, as they may be confused with Classes 1 and 3. On the other hand, Class 5 shows high accuracy classification values, probably due to the fact of having most of the instances from the dataset.

Data augmentation technique used in this project (summarization) has slightly increased classification rates. Even though RoBERTaESP model does not improve the overall accuracy value, most of F1-values in Polarity task (Classes 1, 2 and 3), get higher rates.

Regarding the different datasets created, when SVM model trains on Dataset 2, it classifies correctly much more instances in Class 1 comparing with training with Dataset 1.

Secondly, if Classes 2 and 4 were omitted, SVM models would have been benefited because they have trouble to classify instances in the intermediate classes. Finally, Transformers models improve the results from the SVM models, getting to classify instances into more various classes. Moreover, if Transformers models are compared with each other, RoBERTaESP offers the best accuracy, when training in both Datasets 3 and 4. However, when RoBERTaESP trains with Dataset 4, it gets better F1 scores in every class.

All in all, training with both Title and Opinion fields and also Data Augmentation slightly improves results. Moreover, RoBERTaESP model offers the best results among other Transformers models tested.

Polarity classification on validation dataset							
Model	**Dataset**	**F1 - class 1**	**F1 - class 2**	**F1 - class 3**	**F1 - class 4**	**F1 - class 5**	**Accuracy**
SVM	1	0.04	0.0	0.25	0.0	0.84	0.71
RoBERTaESP	1	0.50	0.38	0.48	0.50	**0.90**	0.78
GPT2	1	0.44	0.36	0.45	0.48	0.88	0.76
SVM	2	0.42	0.02	0.60	0.00	0.84	0.68
RoBERTaESP	2	0.61	0.39	0.64	0.36	0.89	0.74
SVM	3	0.20	0.0	0.32	0.01	0.85	0.72
RoBERTaESP	3	0.57	0.34	0.55	0.53	**0.90**	0.79
SVM	4	0.43	0.01	0.60	0.01	0.84	0.67
RoBERTaESP	4	**0.79**	**0.64**	**0.76**	**0.53**	0.89	**0.79**

Table 6
Results on Polarity task on validation dataset

3.2. Selected model

From the previous results, we can see that the best model is the RoBERTaESP trained using the Dataset 4 (that is, combining title and opinion and using data augmentation). We use this model to predict the outputs for the test dataset provided by the organizers during the evaluation phase which included 12938 instances.

Tables 7 and 8 show our final results in the classification of attractions and polarities, respectively. These results have been published by the organizers of the REST-MEX@IberLef 2022 shared task. Our approach obtains the second-best result from the competition.

The results obtained in the competition closely follow the performance of the selected model with the labeled datasets, as shown in section 3.1.

In the Polarity task, Classes 1, 3 and 4, follow closely 50%, while Class 5 almost reaches 90%. This means that high rated opinions will be classified correctly, while the model may have problems in classifying negative opinions.

The difficulties in predicting the sentiment of an opinion are also due to the number of labels. This is due to the scarce difference between the levels of polarity next to each other. Literature sentiment analysis models perform better on three labels: negative, neutral and positive. If more labels are used then an off by one accuracy is performed [34].

Results from the Attraction task are similar to the performance observed while training models, reaching high accuracy values.

Attraction type classification					
Model	**Dataset**	**F1 - Hotel class**	**F1 - Restaurant class**	**F1 - Attractive class**	**Accuracy**
RoBERTaESP	4	0.9907	0.9811	0.9932	0.9884

Table 7
Results obtained in the Polarity task for selected model in the REST-MEX competition final dataset

Polarity classification							
Model	**Dataset**	**F1 - class 1**	**F1 - class 2**	**F1 - class 3**	**F1 - class 4**	**F1 - class 5**	**Accuracy**
RoBERTaESP	4	0.5190	0.3428	0.5010	0.4799	0.8829	0.7625

Table 8
Our final results for the classification on the attraction type on the test dataset

4. Conclusions

In this paper, we describe our participation at the sentiment analysis track of the Rest-Mex@IberLef 2022 shared task. Different techniques have been evaluated, from SVM to different pretrained Transformers models. Even though SVM is a traditional technique, it is still capable of solving tasks with high accuracy results. However, Transformers have shown the best results using transfer learning.

The best results were obtained with RoBERTaESP, a pretrained transformer model for Spanish language based on RoBERTa. This model obtained an accuracy of 76.25% in the Polarity task and 98.84% in the Attraction task in the competition, being the second-best result in the overall task.

The main difficulties for this task were the unbalanced labeled dataset. In order to solve this problem, we use data augmentation by applying text summarization to increase the number of instances of the classes with fewer instances. This idea proved to bring slight improvements, obtaining more balanced results.

As future work, the task solved in this article could be extended to solve the same problem applied to multiple languages since sentiment analysis is multilingual in nature. Moreover, we plan to explore multi modal approaches to also exploit other input data such as image or sound.

Our experimentation can be found in a Github repository [35].

References

[1] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies 5 (2012) 1–167.

[2] R. Guerrero-Rodriguez, M. Á. 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 (2021) 1–16. doi:https://doi.org/10.1080/13683500.2021.2007227.

[3] C. S. G. López, K. S. B. Arreola, Impacts of tourism and the generation of employment in mexico, Journal of Tourism Analysis: Revista de Análisis Turístico (2019).

[4] A. M. López, Covid-19 in mexico: Impact on tourism 2020, 2021. URL: https://www.statista.com/statistics/1124221/coronavirus-impact-tourism-sector-mexico/.

[5] M. A. Álvarez-Carmona, R. Aranda, A. Y. Rodrıguez-González, L. Pellegrin, H. Carlos, Classifying the mexican epidemiological semaphore colour from the covid-19 text spanish news, Journal of Information Sciences (2022). doi:https://doi.org/10.1177/01655515221100952.

[6] Over a billion reviews and contributions for hotels, attractions, restaurants, and more, ???? URL: https://www.tripadvisor.com/.

[7] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, H. Liu, Natural language processing (nlp) in management research: A literature review, Journal of Management Analytics 7 (2020) 139–172.

[8] M. T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, K. H. Khan, Sentiment analysis and the complex natural language, Complex Adaptive Systems Modeling 4 (2016) 1–19.

[9] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, Advances in neural information processing systems 32 (2019).

[10] C. Colón-Ruiz, I. Segura-Bedmar, Comparing deep learning architectures for sentiment analysis on drug reviews, Journal of Biomedical Informatics 110 (2020) 103539. doi:https://doi.org/10.1016/j.jbi.2020.103539.

[11] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams engineering journal 5 (2014) 1093–1113. doi:https://doi.org/10.1016/j.asej.2014.04.011.

[12] T. Mullen, N. Collier, Sentiment analysis using support vector machines with diverse information sources, in: Proceedings of the 2004 conference on empirical methods in natural language processing, 2004, pp. 412–418.

[13] N. Zainuddin, A. Selamat, Sentiment analysis using support vector machine, in: 2014 international conference on computer, communications, and control technology (I4CT), IEEE, 2014, pp. 333–337.

[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.

[15] R. Socher, Y. Bengio, C. D. Manning, Deep learning for nlp (without magic), in: Tutorial Abstracts of ACL 2012, 2012, pp. 5–5.
[16] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.
[17] 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).
[18] 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).
[19] 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.
[20] C. Shorten, T. M. Khoshgoftaar, B. Furht, Text data augmentation for deep learning, Journal of big Data 8 (2021) 1–34. doi:https://doi.org/10.1186/s40537-021-00492-0.
[21] S. Gupta, S. Gupta, Abstractive summarization: An overview of the state of the art, Expert Systems with Applications 121 (2019) 49–65.
[22] T. Hasan, A. Bhattacharjee, M. S. Islam, K. Mubasshir, Y.-F. Li, Y.-B. Kang, M. S. Rahman, R. Shahriyar, XL-sum: Large-scale multilingual abstractive summarization for 44 languages, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 4693–4703. URL: https://aclanthology.org/2021.findings-acl.413.
[23] C. Aone, M. E. Okurowski, J. Gorlinsky, B. Larsen, A scalable summarization system using robust nlp, in: Intelligent Scalable Text Summarization, 1997.
[24] C. Cortes, V. Vapnik, Support vector machine, Machine learning 20 (1995) 273–297.
[25] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing 408 (2020) 189–215. doi:https://doi.org/10.1016/j.neucom.2019.10.118.
[26] Dccuchile, Dccuchile/spanish-word-embeddings: Spanish word embeddings computed with different methods and from different corpora, ???? URL: https://github.com/dccuchile/spanish-word-embeddings.
[27] J. Canete, Compilation of large spanish unannotated corpora, 2020.
[28] Hugging face – the ai community building the future., ???? URL: https://huggingface.co/.
[29] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, Proceedings of the IEEE 109 (2020) 43–76. doi:https://doi.org/10.1109/JPROC.2020.3004555.
[30] 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.

[31] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, C. Armentano-Oller, C. Rodriguez-Penagos, A. Gonzalez-Agirre, M. Villegas, Maria: Spanish language models, Procesamiento del Lenguaje Natural 68 (2022) 39–60. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6405.

[32] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language models are unsupervised multitask learners (2019).

[33] A. Gillioz, J. Casas, E. Mugellini, O. Abou Khaled, Overview of the transformer-based models for nlp tasks, in: 2020 15th Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2020, pp. 179–183.

[34] C. Delestre, A. Amar, DistilCamemBERT : une distillation du modèle français CamemBERT, in: CAp (Conférence sur l'Apprentissage automatique), Vannes, France, 2022. URL: https://hal.archives-ouvertes.fr/hal-03674695.

[35] Javilonso, Javilonso/nlp_sentiment_uc3m_restmex2022: Https://sites.google.com/cicese.edu.mx/rest-mex-2022/tracks/sentiment-analysis-task, ???? URL: https://github.com/javilonso/NLP_Sentiment_UC3M_RestMex2022.