BERTimbau in Action: An Investigation of its Abilities in Sentiment Analysis, Aspect Extraction, Hate Speech Detection, and Irony Detection.

Júlia da Rocha Junqueira, Félix Leonel V. da Silva, Wesley Costa, Rodrigo B. Carvalho, Alexandre T. Bender, Ulisses B. Corrêa, Larissa A. de Freitas

CDTEC, Universidade Federal de Pelotas
96010-610 - Brazil
{julia.rjunqueira, flvdsilva, wcsilveira, rbcarvalho, atbender, ulisses, larissa}@inf.ufpel.edu.br

Abstract

Social Media has revolutionized how individuals, groups, and communities interact. This immense quantity of unstructured data holds valuable information expressed in informal language. However, automatically extracting this information using Natural Language Processing requires adaptations of traditional methods or the development of new strategies capable of extracting information tackling web-prone language. BERT, a Deep Learning methodology proposed by Google in 2018, brought transfer learning to Natural Language Processing. In this work, we used a BERT model for the Portuguese language called BERTimbau to create models for Sentiment Analysis, Aspect Extraction, Hate Speech Detection, and Irony Detection. We experimented with the two BERTimbau models, base and large. Finally, we compared the results obtained in each task. Experiments with BERTimbau based models obtained improved results, F-Measure of 0.88 and 0.89 in Sentiment Analysis and Hate Speech Detection tasks, respectively, compared to classical Machine Learning approaches.

Introduction

Social Media (SM) has revolutionized the way we live, work, and interact with each other. Their constant evolution has profoundly impacted our culture, reshaping our values, beliefs, and behaviors. The rapid growth and unpredictable nature of SM have propelled us into a new era of connectivity, transforming how we engage with the world and each other. The development of SM platforms can be traced back to the launch of Six Degrees[1] in 1997, which allowed users to create profiles and connect with friends. Other similar SM platforms followed, including Myspace, Facebook, and Twitter.

Since then, other social platforms have emerged, including TikTok, Snapchat, and LinkedIn, each with unique features and a target audience. As SM continues to evolve, we expect to see further innovation and changes in the future, impacting how people connect and communicate.

The SM landscape is comprised of multiple platforms, each with unique characteristics and audiences. Despite their differences, these platforms share a common purpose: to enable people to connect in interactive ways. SM has profoundly impacted communication and information consumption, with positive and negative outcomes. For example, SM has facilitated global connections and communities and has played a role in driving innovation and social change. However, SM has also been associated with cyberbullying, spreading false information, and addiction. While SM presents opportunities and challenges, its impact on modern society must be addressed. As such, it remains a vital aspect of contemporary communication and culture.

Over the past 25 years, Natural Language Processing (NLP) has been extensively researched in traditional media such as newspapers, radio, and television (Farzindar and Inkpen, 2014). With the rise of SM, the need for NLP methods has extended to this new platform. However, the need for Portuguese language corpora limits NLP research in this language. The need for more sufficient data makes developing and testing NLP techniques challenging, reducing the results' accuracy. Therefore, it is essential to address this issue by increasing the size and diversity of Portuguese language corpora to enable more extensive NLP research and practical applications in SM. By better understanding SM communication patterns and Sentiment Analysis, researchers can contribute to developing more efficient and accurate NLP techniques in Portuguese.

Given this context, in this work, we explore the tasks of Sentiment Analysis (SA), Aspect Extraction (AE), Hate Speech Detection (HS), and Irony Detection (ID) using methods based on Deep Learning (DL) applied in SM. For this, we do our experiments in the Portuguese language through the use of BERTimbau base[2] and large[3] (Souza, Nogueira, and Lotufo, 2020).

The paper is structured in the following sections: **Theoretical Background** covers important concepts regarding domain knowledge about Twitter, as well as the technical information relevant to understand the addressed tasks, and finally, a brief background in DL and language models; **Related Works** reviews relevant works previously published

[1]http://www.sixdegrees.com/

[2]https://huggingface.co/neuralmind/bert-base-portuguese-cased

[3]https://huggingface.co/neuralmind/bert-large-portuguese-cased

in the literature, with a particular focus on studies covering NLP models concerning the Portuguese language; **Methodology** describes the steps taken to perform the experiments, including information about datasets, fine-tuning, and the data flux across tasks; **Experiments** shows the configuration and hyperparameters used to approach each task; **Final Remarks** summarizes the work and briefly discusses potential future studies.

Theoretical Background

Twitter[4] is a source of news and trends widely used by journalists, politicians, athletes, and companies. Because of its character limit, communication tends to be more objective, faster, and to the point. The 140 characters limit was valid on Twitter until November 2017; now, tweets can be up to 280 characters. Twitter is used to share ideas and comment about what is happening in real-time. (Boot et al., 2019).

Several websites allow users to write reviews about specific topics related to their content (e.g., a sold product or the quality of their service). Such reviews are often accompanied by a rating system (e.g., “star” ratings, in which the number of stars, usually from 1 to 5, represents your satisfaction with the product). This kind of data does not have a specified format or size; each platform that utilizes this technology internally defines how many characters the users have to express themselves. TripAdvisor[5] is an online travel information and booking website with reviews about various hotels. A hotel’s popularity index on this website determines its customers’ degree of satisfaction based on their personal opinions

SA comprehends the classification of texts in three different polarities, being classified as positive, negative, or neutral. It is considered a complex task because emotions and intentions in the text could result in misunderstanding of the model (Hoang, Bihorac, and Rouces, 2019). Still, regarding SA, the literature depicts three different data processing approaches in terms of granularity: document, sentence, and aspect. The Document-level analysis consists of considering the whole text. The Sentence-level analysis splits the whole text into sentences before processing them individually. The Aspect-level analysis demands more steps to acquire results since it requires AE before classifying their sentiments.

AE is a NLP task that involves identifying and extracting aspects or features being discussed in a given text, such as a product review or customer feedback. These aspects are parts or attributes of an entity, such as a product, service, or event. For example, in a hotel review, aspects may include the room cleanliness, the food quality, and the staff friendliness (Liu, 2015).

HS is a language that attacks or denigrates a specific group based on their characteristics, such as their race, ethnicity, or sexual orientation (Nobata et al., 2016). It is problematic because it involves processing text and understanding the context.

The irony concept is usually understood as a linguistic resource used to express the opposite of the literal meaning of an utterance (Cignarella et al., 2018). ID comprehends the binary classification of texts regarding whether they contain or do not contain ironic behavior.

DL has recently seen a surge in popularity as a way to accelerate the solution of certain types of complex computer problems. Transformers is a DL model introduced in 2017 that uses the Self-Attention mechanism, weighing the influence of different parts of the input data. It aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease. Using Self-Attention, it computes its input and output representations without using sequence-aligned Recurrent Neural Networks or Convolution. It is mainly used in NLP (Vaswani et al., 2017).

Bidirectional Encoder Representations from Transformers (BERT) uses a “Masked Language Model” (MLM), the MLM randomly masks some of the tokens from the input, and the aim is to predict the original vocabulary id of the masked word based only on its context. In addition to the MLM, BERT uses a “Next Sentence Prediction” (NSP) to jointly pre-train text-pair representations (Devlin et al., 2019).

There are a few pre-trained models for the Portuguese language: BioBERTpt, BERTaú, Portuguese BERT base cased QA, and BERTimbau. BioBERTpt, a fine-tuned mBERT on clinical and biomedical texts in Portuguese, achieves state-of-the-art on SemClinBr Named-Entity Recognition (NER) task (Schneider et al., 2020). BERTaú, a BERT model trained from scratch on Portuguese texts of the banking domain, achieves state-of-the-art performance on private in-domain Information Retrieval (IR), SA, and NER tasks (Finardi et al., 2021). Portuguese BERT base cased QA, a fine-tuned on SQUAD v1.1 in Portuguese on Question Answer (QA) task (Guillou, 2021). BERTimbau is a model for Brazilian Portuguese from NeuralmindAI, which achieves state-of-the-art performance in Sentence Textual Similarity (STS), Recognizing Textual Entailment (RTE), and NER (Souza, Nogueira, and Lotufo, 2020).

Related Works

In the literature, we find some works about SA (Document and Aspect-level) in the Portuguese language; they are Barros (2021), Lopes, Corrêa, and Freitas (2021), Lopes et al. (2022), da Silva et al. (2022), Gomes et al. (2022). Also, we find works about HS, as: Pelle and Moreira (2017), Fortuna et al. (2019), Leite et al. (2020), Silva and Freitas (2022) and, we find works about ID, as: Corrêa et al. (2021), Subies (2021), Jiang et al. (2021).

Barros (2021) proposed a new methodology to classify sentiments in texts based on the BERTimbau base and focus on emojis, treating them as an essential source of sentiment instead of considering them as simple input tokens. In addition to giving this focus to emojis, the author also performs a post-training stage with 89,458 samples of texts extracted from SM related to the domain of TV shows, the same domain as the analyzed dataset TweetSentBR (Brum and das Graças Volpe Nunes, 2018). Based on the proposed methodology, the author produces results that surpass state-of-the-art for the dataset TweetSentBR, in which 0.77 Accuracy and an F-Measure of 0.76 were obtained.

[4]https://twitter.com

[5]https://www.tripadvisor.com.br/

Recent studies have applied BERT ‘base-multilingual-cased’ and ‘base-portuguese-cased’ models to perform AE and Aspect Sentiment Classification (ASC) tasks in the Portuguese language (Lopes, Corrêa, and Freitas, 2021; Lopes et al., 2022). These studies have shown promising results in improving the Accuracy of SA of Portuguese texts. In 2022, the Aspect-Based Sentiment Analysis in Portuguese (ABSAPT) competition was introduced as the first shared task dedicated to ABSA in Portuguese texts (da Silva et al., 2022). The competition aims to encourage the development of innovative techniques to address the challenges of ABSA in the Portuguese language.

At the ABSAPT, two sub-tasks were available: AE and ASC. The first task comprehends the identification of aspects of the reviews, and the second task proposes to extract the sentiment orientation (polarity) of the review about a single aspect mentioned in it. Gomes et al. (2022) opted for a more refined approach, using several different models for AE and ASC. The authors obtained the best results using the mDeBERTa base and achieved an Accuracy of 0.67. This paper uses a similar approach to those presented in ABSAPT.

Leite et al. (2020) created a dataset called ToLD-BR. The dataset is divided into 80% of the data for training, 10% of data for development, and 10% of the data for testing. They used Bag-of-Words (BoW) to represent the examples and an AutoML model to build the baseline model (BoW + AutoML); for this, they used the auto-sklearn library for the BERT-based models. They used the simple transformers library, which allows easy training and evaluation. They used default arguments for parameter tuning and defined a seed to allow for reproducibility; two versions of BERT were used BERTimbau base and mBERT. The results obtained were an F-Measure of 0.74 for the BoW + AutoML, 0.76 for the BERTimbau base, and 0.75 for the mBERT.

Silva and Freitas (2022) used the datasets created by (Fortuna et al., 2019) and (Pelle and Moreira, 2017), they used BERTimbau base as a model. The datasets were divided into 80% for training, 10% for validation, and 10% for testing, and some preprocessing was done on the datasets. Three types of data augmentation were used to balance the datasets. They used the BERT default hyperparameters, achieving the best results with oversampling with an F-Measure of 0.91.

In 2021 the IDPT (Irony Detection in Portuguese) competition was proposed (Corrêa et al., 2021). This event was the first shared effort dedicated to identifying the presence of irony in text such as tweets or news in Brazilian Portuguese. Jiang et al. (2021) proposed tackling the problem using BERT, paired with weightloss and ensemble learning. According to the author, the strategy adopted with the best result demonstrates the common usage of the two datasets used in IDPT to assist in model classification and generalization. As a result, the strategy returned a total of 0.48 in Balanced Accuracy. As the IDPT dataset is relatively small, (Subies, 2021) decided to run Data Augmentation techniques. The authors randomly masked 15% of the tokens and used BERTimbau base with hyperparameters GridSearch to predict them. This paper uses similar approaches to those presented in IDPT.

Methodology

Our work is composed of four main steps. Initially, the BERTimbau models (base and large) are used. After, we applied fine-tuning in the SA, AE, HS, and ID tasks. And we test in datasets TweetSentBR (Brum and das Graças Volpe Nunes, 2018), ABSAPT 2022 (da Silva et al., 2022), ToldBR (Leite et al., 2020), and IDPT 2021 (Corrêa et al., 2021). Finally, we analyzed the results obtained in each task.

The main advantage of BERT-based models is that they can be used in different NLP tasks once pre-trained. For this, the fine-tuning task must be carried out, which applies specific training data to the observed problem (Araújo et al., 2022).

BERTimbau (Souza, Nogueira, and Lotufo, 2020), like BERT (Devlin et al., 2019), is a model that accepts sequences of up to 512 tokens as input and produces arrays of 768 or 1024 positions as output. The fine-tuning technique involves adding new layers and training with unseen data to adapt the original model to a specific problem (Araújo et al., 2022). Each dataset used in the tests is divided differently, given that they have different lengths and text contents.

TweetSentBR (Brum and das Graças Volpe Nunes, 2018) is a dataset for Brazilian Portuguese used for the SA task. It contains 15,000 sentences that have TV shows as a domain. The sentences were labeled in three classes (positive, negative, and neutral) by seven annotators from different areas, such as linguistics, journalism, and informatics. The dataset was divided into 6,648 positive sentences, 3,926 neutral sentences, and 4,426 negative sentences. They were extracted from Twitter between January and July 2017.

The dataset used for the AE task is a compilation of reviews from TripAdvisor, prepared by de Freitas (2015) and Corrêa (2021). The training dataset consists of 847 reviews, containing 77 aspects and 3,111 sentiment polarity annotations: 2,112 positive, 472 neutral, and 527 negative examples. The test dataset comprises 184 reviews, having 70 aspects and 686 sentiment polarity annotations, from which 450 are positive, 105 are neutral, and 131 are negative.

ToLD-BR (Leite et al., 2020) is a dataset for Brazilian Portuguese used for the HS task. It contains tweets collected between July and August 2019 using the tool GATE Cloud’s Twitter Collector[6]. They used two strategies to collect the tweets. The first strategy was to search for keywords and predefined hashtags such as *“gay”*, *“little woman”*, and *“northeasterner”*. The second strategy was to collect tweets that mention influential people like the president of Brazil Jair Bolsonaro, and the soccer player Neymar Jr. For the annotation process, 42 annotators classified 1,500 tweets LGBTQ+phobia, obscene, insult, racism, misogyny, or xenophobia. Finally, we obtained 9,245 offensive and 11,693 non-offensive tweets, where three annotators classified each tweet.

The dataset used for the ID task was created to help the IDPT 2021 competition (Corrêa et al., 2021). The training dataset comprises Brazilian Portuguese tweets and news,

[6]https://cloud.gate.ac.uk/info/help/twitter-collector.html

including irony/sarcasm. Tweets was divided by 12,736 ironic and 2,476 non-ironic. The testing dataset contains 300 tweets divided into 177 ironic and 123 non-ironic.

Experiments

For most experiments we used batch size of 32, 4 epochs, learning rate of $2 * 10^{-5}$, loss function CrossEntropy and optimizer AdamW.

The model BERTimbau (base and large) was tested on four tasks SA, AE, ID, and HS. Each test dataset was evaluated on several metrics, such as: Accuracy (Acc), Precision, Recall, F-Measure (F1), and Balanced Accuracy (BAcc) (Brownlee, 2016).

Table 1: Results Obtained.

	Task	Dataset	Acc	Precision	Recall	F1	BAcc
Base	**SA**	TweetSentBR	**0.88**	0.89	0.87	**0.88**	0.88
	AE	ABSAPT 2022	0.26	0.21	0.26	0.19	0.09
	HS	ToLD-BR	**0.88**	0.89	0.88	**0.88**	0.89
	ID	IDPT 2021	0.41	0.36	0.41	0.25	0.49
Large	**SA**	TweetSentBR	**0.89**	0.89	0.89	**0.89**	0.89
	AE	ABSAPT 2022	0.27	0.23	0.27	0.22	0.15
	HS	ToLD-BR	**0.89**	0.90	0.89	**0.89**	0.90
	ID	IDPT 2021	0.40	0.16	0.40	0.22	0.50

In Table 1, we show the results for our experiments with each task for model BERTimbau base and large, respectively. The best results with the BERTimbau base were 0.88 of F-Measure and Accuracy for SA and HS. Still, the best results with the BERTimbau large were 0.89 of F-Measure and Accuracy for the same tasks.

For the SA task, we obtained 0.89 on the F-Measure for the TweetSentBR dataset. Compared to the work of (Barros, 2021), we obtained an increase of 13 percentage points. However, our work used the binary classification, while the Barros (2021) used the ternary classification . In the SA task, we obtained 118 false positives and 112 false negatives. By observing the false positives and false negatives, we can identify possible labeling errors.

From this, a survey was made of how many possible labeling errors could be identified in those sentences in which the model obtained false negatives or false positives. It was identified that of the 118 sentences classified as false positives, in 33 of them, the labeling could be positive. The same task was performed for false negatives, and from 112 sentences classified as false negatives, the labeling could be negatives in 19 sentences.

During the experiments in the AE task, we filter out aspects that have less than 20 occurrences in an attempt to select the meaningful aspects. After this, a total of 25 aspects remain in the dataset. To perform the fine-tuning, we reduce the batch size because of technical limitations. Reviews are usually more extensive than tweets regarding the number of characters. Thus this task is more processing-intensive and requires more memory.

Notably, the experiment results in evidence of a significant imbalance of the dataset used in the task. The major challenge for this area is mainly because AE usually consists of a multi-class classification problem. This further denotes the need for a balanced dataset to approach this task. Even after filtering less common aspects, the high dimensionality of classes corroborates the complexity of the task. Consider a binary problem where randomly classifying samples would yield a 1/2 chance of correctly predicting them. On the other hand, this particular task in the same circumstances would result in a 1/25 chance of correctly classifying samples.

Based on the fact that, after analyzing the data samples, people tend to comment or express their opinion about generic topics rather than minor and more specific ones causing a misleading in the data collected (e.g., “hotel” instead of “room”, and “room service” instead of “food” or “kitchen”).

In the HS task, the large model results in a little better than the base, with both performing better than the original ToLD-BR paper. The work of Leite et al. (2020) has 0.75 of F-Measure; therefore, there are differences between the settings of BERTimbau which makes an exact comparison between the works difficult. During the experiments, we observed a similarity between the hits and misses of the BERTimbau base and large. Both got it wrong when the tweets were written with abbreviated and misspelled words and got it right when the tweets were spelled correctly.

ID task using tweets is a challenging task to achieve. The model with the hyperparameters set as default to the tasks was overfitting in the training data. We obtained the same results using the BERTimbau large with the same hyperparameters as the BERTimbau base run. The results obtained in the ID were low compared to the SA and HS tasks. However, we achieved better results based on a BAcc of 0.49 than the (Jiang et al., 2021) work, which achieved 0.48. It was possible to observe during the results that the model generally misses some tweets that lack previous context, such as *“Tonet is against individual freedom, a true communist”*, that was labeled as non-ironic but missed the context behind it.

Final Remarks

In this work, we explored the capabilities of the model BERTimbau across four different tasks, including SA, AE, HS, and ID. The best results with the BERTimbau base were 0.88 of F-Measure and Accuracy for SA and HS, and BERTimbau large were 0.89 of F-Measure and Accuracy for the same tasks. In the ID task, our model was overfitting in the training data, and the same problem was observed by Jiang et al. (2021), which the author used BERT with the same IDPT 2021 data (Corrêa et al., 2021).

Future works could include: further expanding the investigation to use additional datasets; a more in-depth analysis of tasks where the achieved performance was below the expected; testing other hyperparameters in a particular task; and addressing the overfitting problem in the tweet’s data. Overall, try to achieve better results in Brazilian Portuguese tasks using BERTimbau.

References

Araújo, R. D.; Setti, M.; Berardi, R. C. G.; Graeml, A. R.; Valença, G.; Santos, R.; de Paiva, E. S.; and Pereira, F. S. 2022. Tópicos especiais em sistemas de informação: Minicursos sbsi 2022. *Sociedade Brasileira de Computação*.

Barros, T. M. 2021. Employing transformers and emoji to perform sentiment classification of social media texts. Master's thesis, Universidade Estadual de Campinas.

Boot, A. B.; Tjong Kim Sang, E.; Dijkstra, K.; and Zwaan, R. A. 2019. How character limit affects language usage in tweets. *Palgrave Communications* 5(1).

Brownlee, J. 2016. *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End*. Machine Learning Mastery.

Brum, H. B., and das Graças Volpe Nunes, M. 2018. Building a sentiment corpus of tweets in brazilian portuguese. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*.

Cignarella, A. T.; Frenda, S.; Basile, V.; Bosco, C.; Patti, V.; and Rosso, P. 2018. Overview of the EVALITA 2018 task on irony detection in italian tweets (ironita). In *Proceedings of the 6th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*.

Corrêa, U. B.; Coelho, L.; Santos, L.; and de Freitas, L. A. 2021. Overview of the idpt task on irony detection in portuguese at iberlef 2021. *Procesamiento del Lenguaje Natural* 67.

Corrêa, U. B. 2021. *Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa*. Ph.D. Dissertation, Universidade Federal de Pelotas.

da Silva, F. L. V.; da S. Xavier, G.; Mensenburg, H. M.; Rodrigues, R. F.; dos Santos, L. P.; Araújo, R. M.; Corrêa, U. B.; and de Freitas, L. A. 2022. Absapt 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese. *Procesamiento del Lenguaje Natural* 69.

de Freitas, L. A. 2015. *Feature-level sentiment analysis applied to brazilian portuguese reviews*. Ph.D. Dissertation, Pontifícia Universidade Católica do Rio Grande do Sul.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2019. 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*.

Farzindar, A., and Inkpen, D. 2014. *Natural Language Processing for Social Media*. Springer.

Finardi, P.; Viegas, J. D.; Ferreira, G. T.; , A. F. M.; and Caridá, V. F. 2021. Bertaú: Itaú bert for digital customer service. In *arXiv*.

Fortuna, P.; da Silva, J. R.; Soler-Company, J.; Wanner, L.; and Nunes, S. 2019. A hierarchically-labeled portuguese hate speech dataset. In *Proceedings of the 3rd Workshop on Abusive Language Online*.

Gomes, J. R. S.; Garcia, E. A. S.; Junior, A. F. B.; Rodrigues, R. C.; Silva, D. F. C.; Maia, D. F.; da Silva, N. F. F.; Filho, A. R. G.; and da Silva Soares, A. 2022. Deep learning brasil at ABSAPT 2022: Portuguese transformer ensemble approaches. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR. org*.

Guillou, P. 2021. Portuguese bert base cased qa (question answering), finetuned on squad v1.1.

Hoang, M.; Bihorac, O. A.; and Rouces, J. 2019. Aspect-based sentiment analysis using bert. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*.

Jiang, S.; Chen, C.; Lin, N.; Chen, Z.; and Chen, J. 2021. Irony detection in the portuguese language using bert. *Proceedings http://ceur-ws. org ISSN* 1613.

Leite, J. A.; Silva, D. F.; Bontcheva, K.; and Scarton, C. 2020. Toxic language detection in social media for brazilian portuguese: New dataset and multilingual analysis. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*.

Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Lopes, E. P.; de Freitas, L. A.; Gomes, G.; Lemos, G.; Hammes, L. O.; and Corrêa, U. B. 2022. Exploring bert for aspect-based sentiment analysis in portuguese language. *The International FLAIRS Conference Proceedings* 35.

Lopes, E. P.; Corrêa, U. B.; and Freitas, L. A. 2021. Exploring bert for aspect extraction in portuguese language. *The International FLAIRS Conference Proceedings* 34.

Nobata, C.; Tetreault, J. R.; Thomas, A. O.; Mehdad, Y.; and Chang, Y. 2016. Abusive language detection in online user content. *Proceedings of the 25th International Conference on World Wide Web*.

Pelle, R., and Moreira, V. 2017. Offensive comments in the brazilian web: a dataset and baseline results. In *Proceedings of the 6th Brazilian Workshop on Social Network Analysis and Mining*.

Schneider, E. T. R.; de Souza, J. V. A.; Knafou, J.; Oliveira, L. E. S.; Copara, J.; Gumiel, Y. B.; de Oliveira, L. F. A.; Paraiso, E. C.; Teodoro, D.; and Barra, C. M. C. M. 2020. BioBERTpt - a Portuguese neural language model for clinical named entity recognition. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*.

Silva, F. L. V., and Freitas, L. A. 2022. Brazilian portuguese hate speech classification using bertimbau. *The International FLAIRS Conference Proceedings* 35.

Souza, F.; Nogueira, R.; and Lotufo, R. 2020. Bertimbau: pretrained bert models for brazilian portuguese. In *Proceedings of the 9th Brazilian Conference on Intelligent Systems*.

Subies, G. G. 2021. Guillemgsubies at idpt2021: Identifying irony in portuguese with bert. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2021), co-located with the 37th Conference of the Spanish Society for Natural Language Processing (SEPLN 2021), Online. CEUR. org*, 910–916.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need.