

# Sentiment Analysis in Portuguese Texts from Online Health Community Forums: Data, Model and Evaluation

Yohan Bonescki Gumiel<sup>1,2</sup>, Isabela Lee<sup>3</sup>, Tayane Arantes Soares<sup>3</sup>,  
Thiago Castro Ferreira<sup>1</sup>, Adriana Pagano<sup>3</sup>

<sup>1</sup>Departamento de Ciência da Computação, ICEX,  
Universidade Federal de Minas Gerais  
Belo Horizonte, MG – Brazil

<sup>2</sup>Pontifícia Universidade Católica do Paraná  
Curitiba, PR – Brazil

<sup>3</sup>Faculdade de Letras,  
Universidade Federal de Minas Gerais  
Belo Horizonte, MG – Brazil

yohan.gumiel@pucpr.br

{isabelalee19, arantessoares}@gmail.com

{thiagocf05, apagano}@ufmg.br

**Abstract.** *This study introduces novel data and models for the task of Sentiment Analysis in Portuguese texts about Diabetes Mellitus. The corpus contains 1290 posts retrieved from online health community forums in Portuguese and annotated by two annotators according to 3 sentiment categories (e.g. Positive, Neutral and Negative). Evaluation of traditional (Support Vector Machine, Decision Tree, Random Forest and Logistic Regression classifiers) and state-of-the-art (BERT-based models) machine learning classifiers for the task showed the advantage in performance of the latter models as expected. Data and models are available to the community upon request.*

**Resumo.** *Este estudo apresenta dados e modelos para a Análise de Sentimentos de textos em português sobre Diabetes Mellitus. O corpus é composto por 1290 posts, extraídos de forums online sobre tópicos de saúde e anotados por dois estudantes de acordo com 3 categorias (e.g. Positivo, Neutro e Negativo). A avaliação de classificadores de Aprendizagem de Máquina (classificadores Support Vector Machine, Decision Tree, Random Forest e Logistic Regression) tradicionais e estado-da-arte (modelos baseados em BERT) mostrou a vantagem em performance do segundo tipo como esperado. Os dados e modelos estão disponíveis para a comunidade por meio de solicitação.*

## 1. Introduction

Social media provide a high volume of opinionated data from several sources, such as social networks, forums, and user reviews [Liu 2012]. Among Natural Language Processing (NLP) research concerns, little attention was paid to sentiment analysis until the

early 2000s [Yue et al. 2019]. The growth of social media concurs with advances in sentiment analysis research, this topic having become central in social media-related research nowadays [Liu 2012, Yadav and Vishwakarma 2020]. Sentiment analysis benefits from opinionated data to analyze the sentiments (attitudes, feelings, and opinions) towards entities [Yadav and Vishwakarma 2020]. Among a myriad of social media texts there are forums dedicated to particular topics regarding healthcare, which have become increasingly popular in recent years and which are worth exploring as sources of insight for strategies to reduce inequalities and promote well-being, two relevant goals in the UN 2030 Agenda for Sustainable Development.

People may not have ready access to primary care centers for preventive exams or they may find health professionals are not always available to provide adequate and detailed answers to all patients' questions or provide guidance to the patient and their families about all essential aspects of patient care during the consultation time [Gabarron et al. 2018]. Thus, patients often use social media to ask health-related questions and share experiences with other users with the same condition. Our study focuses on one such use of social media, namely online open-access communities. In particular, it draws on posts extracted from online health forums on diabetes, an increasingly prevalent severe and long-term condition affecting people worldwide and rapidly growing in Brazil. According to the most recent survey, dated from 2019, it is estimated that almost half a billion people (9.3% of the adults between 20-79 years) were living with diabetes in 2019 [Saeedi et al. 2019]. Most alarmingly, one every two (50.1%) persons living with diabetes is not aware or has not been diagnosed as having diabetes. Hence, studies that aim to explore the language used in social media by people who seek advice on diagnosed or suspected diabetes are expected to contribute to devising healthcare tools for better support and prevention. Our study seeks to develop a framework to automatically detect social media user sentiment modelled upon data from diabetes online health forum posts. This framework is scheduled to be integrated with other applications, such as chatbots, to control the bot dialog flow according to the user's sentiment.

Prior studies analyzed online health communities' posts, tweets, and news-related comments for the diabetes context. [Gabarron et al. 2019, Salas-Zárate et al. 2017, Cignarelli et al. 2020] analyzed diabetes-related tweets, whereas [Liu et al. 2020, Lu et al. 2017] drew on online health communities' posts. Further, the study by [Liu et al. 2018] examined opinion news about diabetes and their respective comments. None of these studies, however, has addressed social media posts in online communities in Brazil. There is a lack of social media datasets and studies for sentiment analysis for the Portuguese language considering the healthcare domain. Hence, we purported to fill this gap by annotating a sentiment analysis dataset and evaluating the sentiment extraction results using traditional and state-of-the-art machine learning classifiers.

**Related Work** The study by [Liu et al. 2018] noticed that positive emotions were expressed more frequently than negative sentiments when analyzing the sentiment over an online health community about diabetes. However, the posts collected for our corpus had a different characteristic, being generally more negative or neutral.

	Sentiment	Post
#1	Positive	Pessoal, obrigado para quem orou e para quem fez oração. Meu médico me disse que não sabe como, mas eu fiz meus exames e está tudo ótimo. Graças a Deus. Tirou até a insulina e vou tomar apenas um Glifage. Obrigada, Deus. EMOJIS <i>Folks, thank you to those of you who prayed for me. My doctor told me he's got no explanation for this, I got my test results and everything is fine. Thank God. He even took insulin out of my prescription and I'm only taking a Glifage pill. Thank you, Lord. EMOJIS</i>
#2	Neutral	O que vocês acham dos derivados do leite? Por exemplo: queijo, manteiga, iogurte. Faz bem ou mal para a saúde do diabético? <i>What do you think about dairy products? For example: cheese, butter, yoghurt. Are they good or bad for diabetic people's health?</i>
#3	Negative	Estou com minha unha do dedão esbranquiçada e dolorida. Está muito fraca e quebradiça. Alguém sabe de algo que posso usar? <i>My toenail is whitish and sore. It is very weak and brittle. Can anyone suggest anything for me to try?</i>
#4	Neutral	Podemos comer amendoim japoneses? <i>Can we eat Japanese-style peanuts?</i>
#5	Neutral	Posso fazer o suco de chuchu, sem limão? <i>Can I prepare chayote juice with no lemon?</i>
#6	Negative	Sensação doce na boca, mesmo sem comer doces. Por que será? <i>Sweet aftertaste in my mouth, even when I haven't had any sweets. How come?</i>

**Table 1. Examples of posts with annotated sentiments and free-translation into English**

## 2. Corpus

**Material** We collected a total of 1290 texts from open Social Media forums on diabetes management and self-care in Brazil, encompassing communities with over 80 thousand Portuguese speaking users.

**Annotators** Annotation of the texts was performed by two undergraduate students in Language and Arts.

**Anonymization** To ensure users' privacy and anonymity, before the annotation starts, the selected texts were de-identified by first removing some emojis, fixing orthographic mistakes and paraphrasing non-fluent syntactic structures. Then any identifier, such as name, phone or address, was removed from the questions. Moreover, *quasi-identifiers*, such as age and relative mentions, were modified. Users' age were modified by randomly choosing a number in the interval of  $[age - 5; age + 5]$ , whereas mentions to relatives were randomly changed by the reference to a relative with similar age, such as *parent*  $\leftrightarrow$  *uncle*, *parent in law*; *sibling*  $\leftrightarrow$  *cousin*, *partner*; *son*  $\leftrightarrow$  *nephew*.

**Annotation** The texts were annotated with sentiment polarity according to three categories: Positive, Negative and Neutral. The annotations were conducted with Label Studio, an open-source web annotation tool [Tkachenko et al. 2021].

The annotators were requested to consider the overall sentiment polarity in each post bearing in mind semantic and pragmatic features of the texts. Exclamation marks were considered intensification of emotion, leading to texts that contained those punctuation marks being annotated as positive or negative. As the domain of the texts pertains to a chronic health condition, negative polarity was expected to be dominant. In view of this, a more fine-grained definition of polarity categories was needed so that features indicating positive and neutral polarity could be detected in the analysis. Hence the 3 polarity categories were defined as:

- *Positive*: statements by users in the first person with references to their experience regarding personal achievement and counseling others towards adoption of healthy habits and promotion of well-being
- *Negative*: statements by users in the first person with references to their experience regarding symptoms and difficulties in managing diabetes
- *Neutral*: objective statements, with no reference to personal experience, pertaining to seeking general information, frequently on food and medication

By "references to experience", we mean instances in which the user states his/her personal opinion and explicitly describes his/her feelings in terms of physical and mental state.

Examples of the three sentiment categories are illustrated in examples #1, #2, and #3 in Table 1. Some posts demanded closer inspection to annotate their sentiment. For instance, texts with 1st person singular or plural pronominal forms but not construing any personal meaning were classified as Neutral. Examples #4 and #5 in Table 1 illustrate such case: the posts are written by users seeking information on food. Both use a 1st

Sentiment	Annotations	Avg. Characters	Avg. Tokens
Positive	44	156.32	27.57
Neutral	575	71.3	12.35
Negative	671	146.27	25.54
All	1290	113.19	19.73

**Table 2. Corpus Statistics**

Model	Precision	Recall	F1-Score
Decision Tree	63.92 ( $\pm 0.01$ )	65.16 ( $\pm 0.01$ )	64.43 ( $\pm 0.01$ )
Random Forest	74.67 ( $\pm 0.00$ )	76.43 ( $\pm 0.00$ )	75.01 ( $\pm 0.00$ )
Logistic Regression	72.95 ( $\pm 0.00$ )	75.58 ( $\pm 0.00$ )	74.24 ( $\pm 0.00$ )
SVM	70.79 ( $\pm 0.00$ )	73.26 ( $\pm 0.00$ )	71.95 ( $\pm 0.00$ )
mBERT	77.62 ( $\pm 0.02$ )	77.52 ( $\pm 0.02$ )	76.79 ( $\pm 0.02$ )
BioBERTpt	78.56 ( $\pm 0.02$ )	77.29 ( $\pm 0.02$ )	76.23 ( $\pm 0.02$ )
BERTimbau	<b>83.12</b> ( $\pm 0.01$ )	<b>82.67</b> ( $\pm 0.01$ )	<b>82.53</b> ( $\pm 0.01$ )

**Table 3. Classifiers performance considering weighted precision, recall and F1-score, averaged after 10 runs.**

person pronoun, but they do not refer to the users’ personal experiences. Thus, they were labelled Neutral. In contrast, example #6 in Table 1 illustrates a post which was annotated as Negative, even though no first person pronoun is used in Portuguese. The post construes an individual experience linked to a common diabetes symptom, namely a sweet aftertaste, a likely warning sign that the person may have diabetes.

**Inter-annotator Agreement** Using Cohen-kappa [Cohen 1960], we computed an inter-annotator agreement value of 0.64, which can be considered as a substantial agreement by the scale provided by [Landis and Koch 1977]. The annotators took on average 47.71 seconds to classify the sentiment in each text.

**Corpus** Our corpus statistics are shown in Table 2. From our 1290 annotated texts, most of them had Neutral and Negative annotations, with a total of 575 (44.57%) Neutral sentiment annotations and 671 (52.02%) Negative sentiment annotations. There were few Positive sentiment annotations in the corpus, totalling 44 (3.41%) annotations. With regard to average number of characters and tokens, Positive annotations and Negative ones had similar numbers, whereas Neutral annotations had a smaller average number of characters and tokens. This may suggest that Neutral posts tend to be shorter than Positive and Negative ones.

### 3. Models

We evaluated several traditional and state-of-the-art classifiers for sentiment analysis in the domain of Diabetes. As traditional ones, we performed experiments with Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT) and Logistic Regression (LR) classifiers. As input features to the classic classifiers, we used the TF-IDF vector representations of the texts to be classified.

		Predicted		
		Negative	Neutral	Positive
Actual	Negative	123	11	0
	Neutral	32	83	0
	Positive	4	2	3

**Figure 1. Confusion matrix for the BERTimbau model.**

As state-of-the-art approaches, we fine-tuned BERT-based models such as its multilingual version [Devlin et al. 2019]; BERTimbau [Souza et al. 2020], its Brazilian Portuguese version; and BioBERTpt [Schneider et al. 2020], a Brazilian Portuguese version of BERT focused on the clinical domain.

## 4. Experiment

**Data** To evaluate the approaches, our collected corpus was divided into training set, development set and test set. We selected 774 texts (60%) for the training set, 258 (20%) for development set and 258 (20%) for testing set, maintaining the class stratification among the splits.

**Method** The sentiment polarity extraction was treated as a multi-class classification task, where each model was trained to predict one out of the three sentiment categories (e.g., Positive, Neutral or Negative) given the input text. Models were evaluated by computing the weighted Precision, Recall and F-Score in the test set, averaged after 10 runs.

## 5. Results

Results of the model of sentiment analysis in the medical domain are summarized in Table 3. Traditional machine learning classifiers did not perform as well as the contextual models, as expected. From the contextual models, the fine-tuning of BERTimbau achieved the best weighted F1-score, being able to infer correct predictions for both minority and majority classes. BERTimbau was followed by mBERT and BioBERTpt, with both having similar performance, with slightly superior results for mBERT.

Aiming to verify the label distribution over the predictions for our best model (BERTimbau), we provide a confusion matrix in Figure 1. The BERTimbau model was not that effective in predicting Neutral and Positive classes. Examples that had a Neutral annotation were miss-classified as Negative several times. Further, due to the lack of Positive examples, the model incorrectly predicted examples with Positive annotations as Negative and Neutral annotations.

## 6. Discussion

This study introduces novel data and models for the task of Sentiment Analysis in Portuguese texts about Diabetes Mellitus.

**Corpus** We had our texts annotated according to three sentiment categories: Positive, Neutral and Negative. There was a smaller amount of Positive annotations in comparison to Negative and Neutral annotations. In general, users tended to have a negative or neutral sentiment when asking questions or talking about their problems, primarily because of their anxiety or uncertainty about the related topic. According to the study by [Gabarron et al. 2019], which analyzed sentiments over diabetes-related tweets, they noticed that tweets about Type 2 Diabetes were more negative, particularly when there were no emojis. For privacy reasons, many of the emojis in the texts of our corpus were removed and perhaps this may be a bottleneck from our corpus. On the other hand, we judged the anonymization process as necessary in order to respect privacy of the users and in agreement with ethical concerns.

In addition, [Gabarron et al. 2019] noticed that tweets about Type 2 Diabetes were more negative than tweets about Type 1 Diabetes. In our corpus we made no distinction about Diabetes type; however, diabetes type could be an influencing factor over post negativity prevalence.

**Classifiers** We trained both traditional and state-of-the-art machine learning classifiers for the target task. As expected, the current state-of-the-art approaches based on contextual models trained over large corpora achieved overall superior results. Further, the pre-trained models specific to Portuguese (BioBERTpt and BERTimbau) showed improved performance over mBERT, which is pre-trained over several languages. Additionally, the BERTimbau, which is pre-trained over a large and diverse corpus of internet pages, had superior performance compared to BioBERTpt, which is initialized with mBERT and trained on clinical notes and biomedical literature. The fine-tuning of BERTimbau had the f1-scores on micro, macro, and weighted evaluation, showing its ability to predict both majority and minority classes.

## 7. Conclusion

We found diabetes posts in social media a sensitive context, with few positive posts. Hence, predicting positive sentiments became a challenge to the classifiers. However, the state-of-the-art models, especially the fine-tuning of BERTimbau, were able to overcome that challenge.

Contextual models that were specific to the Portuguese language achieved overall superior performance, showing the benefit of developing models that are customized to a particular language.

As future work, we plan to expand our corpus, especially the number of positive annotations. Further, we plan to expand this study by including posts on other chronic conditions or diseases. Additionally, we plan to provide further experiments with traditional machine learning classifiers with ablation tests to verify the effect of other features besides representations from TF-IDF.

## References

Cignarelli, A., Sansone, A., Caruso, I., Perrini, S., Natalicchio, A., Laviola, L., Jannini, E. A., and Giorgino, F. (2020). Diabetes in the time of covid-19: A twitter-based sentiment analysis. *Journal of Diabetes Science and Technology*, 14(6):1131–1132.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Gabarron, E., Bradway, M., Fernandez-Luque, L., Chomutare, T., Hansen, A. H., Wynn, R., and Årsand, E. (2018). Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. *BMC Health Services Research*, 18(1):414.
- Gabarron, E., Dorrnazor, E., Rivera-Romero, O., and Wynn, R. (2019). Diabetes on twitter: A sentiment analysis. *Journal of Diabetes Science and Technology*, 13(3):439–444.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan Claypool Publishers.
- Liu, X., Sun, M., and Li, J. (2018). Research on gender differences in online health communities. *International Journal of Medical Informatics*, 111:172–181.
- Liu, Y., Stouffs, R., and Theng, Y. L. (2020). Sentiment analysis on social media for identifying public awareness of type 2 diabetes. In *The 54th International Conference of the Architectural Science Association (ANZAScA)*.
- Lu, Y., Wu, Y., Liu, J., Li, J., and Zhang, P. (2017). Understanding health care social media use from different stakeholder perspectives: A content analysis of an online health community. *J Med Internet Res*, 19(4):e109.
- Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., and Williams, R. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas, 9th edition. *Diabetes Research and Clinical Practice*, 157:107843.
- Salas-Zárate, M. d. P., Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodríguez-García, M. Á., and Valencia-García, R. (2017). Sentiment analysis on tweets about diabetes: An aspect-level approach. *Computational and Mathematical Methods in Medicine*, 2017:5140631.
- Schneider, E. T. R., de Souza, J. V. A., Knafou, J., Oliveira, L. E. S. e., Copara, J., Gumiel, Y. B., Oliveira, L. F. A. d., 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*, pages 65–72. Association for Computational Linguistics.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: Pretrained bert models for brazilian portuguese. In Cerri, R. and Prati, R. C., editors, *Intelligent Systems*, pages 403–417, Cham. Springer International Publishing.



- Tkachenko, M., Malyuk, M., Shevchenko, N., Holmanyuk, A., and Liubimov, N. (2020-2021). Label Studio: Data labeling software. Open source software available from <https://github.com/heartexlabs/label-studio>.
- Yadav, A. and Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6):4335–4385.
- Yue, L., Chen, W., Li, X., Zuo, W., and Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2):617–663.