- ORIGINAL ARTICLE -

Evaluation of Approaches Based on the BERT Model for Opinion Mining about the Cachaça Beverage

Evaluación de enfoques basados en el modelo BERT para la minería de opinión sobre la bebida Cachaça

Thiago Salles Santos[1], Mozar José Brito[2], and Denilson Alves Pereira[1]

[1]*Department of Computer Science, Federal University of Lavras, Lavras, MG, Brazil*
thiagosallessantos3@gmail.com, denilsonpereira@ufla.br
[2]*Department of Agroindustrial Management, Federal University of Lavras, Lavras, MG, Brazil*
mozarjdb@ufla.br

Abstract

Opinion mining is a natural language processing task that aims to classify user opinions expressed on e-commerce platforms, social networks and other media. It is an important tool for decision making, monitoring products/services, detecting trends, developing marketing strategies, among others. Much research has been carried out addressing opinions in the English language. The Portuguese language is still very lacking in linguistic resources aimed at training machine learning models. This work contributes to the evaluation of approaches based on the BERT language model for opinion mining in the Portuguese language, in particular, by creating and evaluating a dataset with labeled data in the domain of the beverage called Cachaça. This is a popular drink in Brazil, and of great economic importance. As a result of the experimental evaluation, the approaches based on the BERT model stood out in relation to two baselines, and in a cross-domain evaluation, they achieved values greater than 0.97 in the F1 metric for classification into 2 classes and 0.64 for 3 classes, in the dataset labeled for the cachaça beverage.

Keywords: BERT, Cachaça, Opinion Mining, Sentiment Analysis, Social Network

Resumen

La minería de opinión es una tarea de procesamiento del lenguaje natural que busca clasificar las opiniones de los usuarios expresadas en plataformas de comercio electrónico, redes sociales y otros medios. Es una herramienta importante para la toma de decisiones, el monitoreo de productos/servicios, la detección de tendencias y el desarrollo de estrategias de marketing, entre otros. Se han realizado numerosas investigaciones sobre opiniones en inglés. El portugués aún carece de recursos lingüísticos para el entrenamiento de modelos de aprendizaje automático. Este trabajo contribuye a la evaluación de enfoques basados en el modelo lingüístico BERT para la minería de opinión en portugués, en particular mediante la creación y evaluación de un conjunto de datos etiquetados en el dominio de la cachaça. Esta bebida es popular en Brasil y de gran importancia económica. Como resultado de la evaluación experimental, los enfoques basados en el modelo BERT destacaron en relación con dos líneas base y, en una evaluación interdominio, alcanzaron valores superiores a 0,97 en la métrica F1 para la clasificación en dos clases y 0,64 para tres clases, en el conjunto de datos etiquetado para la cachaça.

Palabras claves: Análisis de Sentimientos, BERT, Cachaza, Minería de Opinión, Redes Sociales

1 Introduction

Sentiment analysis is an area of research that aims to find computational solutions to identify and quantify opinions, attitudes and emotions expressed in texts and other media [1]. It is a valuable tool in applications such as brand monitoring, market research, political decision making, trend detection, among others.

Among the subtasks of sentiment analysis, in this work, we deal with opinion mining, which consists of classifying the text as containing a positive, negative or neutral opinion. This is distinct from the emotion mining subtask, which consists of identifying the types of emotions contained in the text, such as happiness, sadness, anger and disgust [2]. Despite the advances achieved, sentiment analysis still presents challenges in interpreting linguistic nuances, such as sarcasm, irony, denial, idiomatic expressions and other ambiguities characteristic of human language. Furthermore, for languages such as Portuguese, the linguistic resources available to be used by Natural Language Processing (NLP) techniques are still much scarcer than for the English language [3].

Sentiment analysis uses knowledge-based techniques, statistical methods, and hybrid approaches. With the advancement of deep learning-based models, such as BERT and multilingual Large Language

Models (LLMs) [4], these techniques have achieved better results. However, in informal texts, the use of culturally typical expressions still poses a challenge to models. For example, the positive user's opinion of a certain beverage "Deu água na boca, essa marvada!" (It gave water in the mouth, this evil!) contains slang that is difficult for language models, which are trained mostly on English data, to interpret.

In this work, we address opinion mining in the specific domain of the Brazilian beverage called Cachaça. It is an alcoholic beverage produced from sugarcane. It is known worldwide for its use in the cocktail Caipirinha. Cachaça production started at the beginning of Portuguese colonization in Brazil. It is a very versatile drink, and can be consumed neat, chilled, or mixed with other beverages [5].

To better understand the opinion of consumers of the cachaça beverage, we collected a set of posts and comments about the drink from the social networks Facebook and X (formerly Twitter). A sample from this dataset was manually labeled into positive, negative, and neutral classes. We implemented classifiers based on the BERT (Bidirectional Encoder Representations from Transformers) language model [6] to evaluate the labeled dataset.

One of the difficulties in classifying data in specific domains is the lack of labeled data to train machine learning models. The cost of labeling is very high, as it requires manual effort from subject matter experts. We did not find labeled datasets in the literature for opinion mining about the cachaça beverage. Therefore, in this work, we evaluate how useful and generalizable data from other domains, annotated for opinion classification, can be to the cachaça domain. We used data available in the works of [7] and [8]. The data from [7] is from different domains of product reviews from e-commerce stores and movies, and the data from [8] are tweets on TV show domain.

Our main contributions are: (i) evaluation of classifiers based on the BERT model for opinion mining; (ii) collection and labeling of a dataset for the cachaça beverage domain and (iii) experimental evaluation of trained classification models in different domains to predict data about the cachaça beverage.

As a result, our approaches based on the BERT model for opinion mining obtained higher values in the ROC-AUC, Accuracy and F1 metrics in most evaluations when compared with two baselines. We collected a dataset with 52,205 comments about the cachaça beverage, extracted from the social networks Facebook and X, of which a sample of 2,200 comments were labeled in the positive, negative and neutral classes. In the cross-domain evaluation, our approaches based on the BERT model achieved values greater than 0.97 in the F1 metric for classification into 2 classes and 0.64 for 3 classes, in the dataset labeled for the cachaça beverage.

The remainder of this document is organized as follows. Section 3 describes the main work related to our research. Section 4 describes the experimental evaluation of approaches based on the BERT language model for the task of opinion mining in Portuguese. Section 5 deals with the process for collecting and labeling our dataset containing comments about the cachaça beverage, obtained from social networks. Section 6 describes the experimental evaluation of our cachaça dataset. And finally, Section 7 presents the research conclusions.

2 Context of Cachaça in Brazil

The history of cachaça production and consumption has been marked by multiple interpretations, such as the writings of [9] that gave rise to the book Prelúdio da Cachaça. The author points out that the production and consumption of cachaça began in the colonial period, around 1532, during the process of installing the first sugar mills along the Brazilian coast.

The consumption of cachaça was widely stigmatized, to the point that its commercialization was banned and resumed under a strong taxation regime that made several cachaça production units unviable. Despite this critical incident, the production and consumption of cachaça underwent a reorganization process that changed the socioeconomic status of this drink. This process was decisive in increasing the social prestige of cachaça and expanding the market for the drink, which began to be consumed by people living in different countries around the world.

The Cachaça Yearbook, published by the Ministry of Agriculture, Livestock and Supply (MAPA)[1], points out that in 2021, Brazil had 936 cachaça production units that were registered. These production units, which add value to cachaça production and consumption, are responsible for offering 5,926 brands. In 2021, domestic consumption of cachaça was 520.9 million liters, representing 72% of the country's distilled beverages market. It is noteworthy that 98% of Brazilian cachaça production has been carried out by small and medium-sized producers. This indicator expresses, in part, the social and economic relevance of the production of this drink whose exports, in 2022, reached the mark of US$18.47 million, the highest value in the last 12 years according to information from the Brazilian government that was released by the Brazilian Cachaça Institute (IBRAC)[2]. The cachaça production chain employs around 600 thousand people who are involved in its production and market process that involves Brazil and 67 other countries.

In addition to generating employment and income, the consumption of cachaça is marked by socially shared cultural traits and the meanings attributed to the

[1] https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/publicacoes/anuario-da-cachaca-2021-1.pdf

[2] https://ibrac.net/

act of consumption of this alcoholic beverage, which has become Brazilian cultural and intangible heritage. The indexation between culture and the production of meanings about the act of drinking alcoholic beverages gives rise to multiple feelings. In other words, drinking cachaça can evoke a series of feelings that can take on a positive or negative character. The act of consuming cachaça can produce feelings of: i) happiness that is associated with cultural rites of celebration and commemoration of special occasions; ii) relaxation resulting from cultural rites that reduce the effects of pressure and anxiety; iii) connections that involve establishing relationships with people; iv) socialization and sharing that allow the construction of new forms of coexistence and the search for solutions to everyday problems; v) joy that involves well-being and the pursuit of pleasure. In addition to these positive feelings, drinking can also evoke negative feelings, such as: i) anxiety resulting from excessive alcohol consumption can lead to anxiety and nervousness; ii) depression that can lead to social isolation; iii) development of behaviors marked by violence due to excessive alcohol consumption; iv) dependence that produces harmful effects on health and the ability to experience life in a positive way. These potential negative effects of drinking can vary from person to person, depending on a variety of factors, such as culture, personality and personal experiences.

3 Related Work

The development of sentiment analysis methods is of great interest to the scientific community, industry and government. A multitude of works have been developed in recent years [10, 11, 12, 13, 14, 15, 2, 16, 17]. However, many challenges still need to be overcome to accurately interpret feelings and determine their polarity [1].

Most works on sentiment analysis are in English. Thus, this language has many more linguistic resources available for analysis, such as sentiment lexicons, ontologies, language models, datasets, and tools. Languages such as Portuguese are much more lacking in these resources [3]. Still, many works have been developed for the Portuguese language [18, 19, 20, 21, 22, 23].

In recent contributions to linguistic resources in the Portuguese language, [22] collected and labeled a dataset with reviews about supermarkets from all over Brazil for the opinion mining task. [23] labeled a stance dataset collected from Twitter for sentiment analysis, and they tested a baseline model using a pre-trained BERT model for Portuguese.

The BERT language model [6] is used in several NLP tasks, including opinion mining [24, 23]. In [24], the authors evaluate different alternatives for using the pre-trained BERT model on corpus in the Portuguese language [25] and in multiple languages. In this work, we also evaluated the BERT model on some of its datasets and found similar results.

This work is part of a project on the development of artificial intelligence resources to improve the cachaça beverage market. One of the linguistic resources already published was the CachacaNER dataset [26], which contains data labeled into 17 categories for the named entity recognition task.

Many works deal with the cachaça market. [27] investigate the institutional practices that reordered the status of still artisanal cachaça in the State of Minas Gerais, Brazil. [28] analyze the construction of the brand identity of cachaça manufacturing companies in the state of Santa Catarina, Brazil. [29] research the consumption of cachaça, seeking to understand the meanings attributed to it by its consumers.

No studies were found in the literature that use automatic sentiment analysis techniques to identify the opinion of cachaça consumers, as we did in this work. We found some research involving cachaça and artificial intelligence. In [30], the authors present a method for automatically recognizing the type of wood used in aging cachaça using information from a computer vision system and two machine learning classifiers. [31] present a machine learning-based approach for classifying barrel-aged cachaça based on maturation level. Related to the sugarcane crop, [32] present a review of studies that implemented several machine learning algorithms based on remote sensing data in mapping and classifying this crop.

[7] carried out various evaluations on datasets labeled for opinion mining, including cross-domain evaluations, seeking to identify generalization capabilities. In this work, we use part of the data organized by them to evaluate our classifiers and train models for prediction on our cachaça dataset. We also use data from [8] for similar evaluations. In addition to collecting and labeling the dataset, called TweetSentBR, the authors evaluated it using six machine learning classifiers.

4 Evaluation of the BERT Model for Opinion Mining

One of the objectives of this work was to evaluate the BERT language model for the opinion mining task in Portuguese. In the following subsections, we discuss details of the experimental evaluation and the results obtained.

4.1 Datasets and Baselines

We used part of the datasets organized by [7] and [8] to evaluate our approach. We also use the results of these two works as a baseline for comparison with the results of our proposal.

The datasets made available[3] by [7] have fixed

[3]https://www.kaggle.com/datasets/fredericods/ptbr-sentiment-analysis-datasets

partitions of data in training, testing and validation, which allows fair comparisons of different algorithms. They were generated from five user review datasets: Olist [33], B2W [34], Buscapé [35], UTLC-Apps and UTLC-Movies [36]. Details of how the data were cleaned and preprocessed are available in [7]. Due to computational resource limitations, we were unable to process the Buscapé dataset. Therefore, it was not used in our studies. We use the polarity datasets, with binary labels (positive and negative). Table 1 summarizes some statistics per dataset.

The TweetSentBR dataset was made available[4] by [8], and belongs to the TV show domain. After preprocessing that removed repeated and unlabeled sentences from the original file, the resulting dataset has the statistics shown in Table 1. This dataset was made available in training and testing partitions and is labeled into three classes: positive, negative and neutral.

The choice of these datasets is due to the fact that, in general, the dataset provides a solid basis for analyzing consumer opinions and preferences in relation to products, in line with our objective of analyzing sentiments in the domain of the cachaça product. The TweetSentBR dataset was included because it is composed of *tweets*, carrying with it the style and behavior of the vocabulary used on social networks, linking to our cachaça datasets, which were also extracted from social networks. Furthermore, these datasets contain emotional expressions that may also appear in the cachaça domain, justifying their use in a cross-domain evaluation.

[7] evaluated their datasets using machine learning classifiers and some document embedding strategies. They also evaluated the generalization capacity of their models in cross-domain contexts, i.e., one model developed under one dataset is evaluated with another one, and vice and versa. [8] also evaluated their dataset using machine learning classifiers.

4.2 Experimental setup

The BERT model [6] is one of the most prominent Transformer-based models, extensively studied and evaluated for NLP problems. In this work, we use two different variants of the BERT model: the BERTimbau Base [25], a Portuguese BERT variant trained with the Brazilian Web as Corpus (BrWaC) [37], and the twitter-XLM-roBERTa [38], which is a model trained with 198 million *tweets* in 8 languages, including Portuguese, and fine-tuned for the opinion mining task. This model is based on the XLM-roBERTa [39] model, which is a multilingual version of the roBERTa [40] model. In other words, we implemented and trained a classifier based on the BERTimbau model, as we will describe in detail below, and used an already trained classifier for opinion mining based on the XLM-roBERTa model. All language models used are available on Hugging Face[5], which is a platform for sharing models and datasets.

Opinion mining is modeled as a text classification problem. Our classifiers employ a model architecture composed of a BERT model, a linear layer on top of the BERT's hidden-states output and a softmax. The linear layer is preceded by a dropout layer and its number of nodes is equal to the number of classes. The BERTimbau-based classifiers were fine-tuned by each training dataset, where all weights, including BERT's, are updated jointly during training. The twitter-XLM-roBERTa, on the other hand, comes pretrained as described in [38]. The authors used their tweets dataset on training, and integrated an adapter technique, by means of which they freeze the language model and only fine-tune one additional classification layer.

The hyperparameters adopted in the fine-tuning process of the BERTimbau model were based on the article [6], with the number of epochs equal to 3, Adam optimizer, learning rate equal to $2e^{-5}$ and batches of size 16 or 32.

The experiments used the classic training-testing hold-out scheme, using the data partitions provided by [7] and [8]. In the same way as in [7], we concatenated the first nine folds for the training set, and evaluated the results in the tenth fold.

To carry out the experiments, we used a computer with an NVIDIA GeForce RTX 3090 graphics card, 24GB VRAM, 128GB of RAM, 2.9GHz Intel Core i7 processor, 10^a generation. The server runs the Ubuntu 22.04.3 LTS operating system and the Python programming language version 3.9.13, in addition to the Tensorflow libraries version 2.7.0, Transformers 4.24.0 and Datasets 2.6.1.

4.3 Evaluation metrics

In the following equations, let TP (true positive) be the number of instances correctly predicted for a specific class, FP (false positive) the number of instances incorrectly predicted, FN (false negative) the number of instances incorrectly not predicted, and TN (true negative) the number of instances correctly not predicted. We use the metrics Precision, calculated as $P = \frac{TP}{TP+FP}$, Recall, calculated as $R = \frac{TP}{TP+FN}$, F1-Score, calculated as $F1 = 2 * \frac{P*R}{P+R}$. The results reported in the experiments are the arithmetic average between all classes (Macro-Average). The Accuracy metric was also used, calculated as $A = \frac{TP+TN}{TP+TN+FP+FN}$.

Furthermore, we used the ROC-AUC metric (Area Under the Receiver Operating Characteristic Curve), which is used to evaluate the performance of binary classification models. It measures the model's ability to correctly discriminate between positive and negative classes at different classification thresholds. It can take values from 0 to 1. A higher ROC-AUC indicates better performance. A perfect model would have an

[4]https://bitbucket.org/HBrum/tweetsentbr/

[5]https://huggingface.co/

Table 1: Number of samples, number of tokens per sample (mean, median, mode, min, max), vocabulary size, and the polarity distribution (positive, negative and neutral) of each dataset

Dataset	#sample	Mean	Median	Mode	Min	Max	Vocab. size	Pos/Neg/Neu labels (%)
Olist	37,953	7	6	2	1	35	14,601	70.0 / 30.0 / 0.0
B2W	115,977	14	10	7	1	618	48,211	69.2 / 30.8 / 0.0
UTLC-Apps	968,018	8	5	1	1	356	140,388	77.5 / 22.5 / 0.0
UTLC-Movies	1,188.497	21	10	2	1	4.515	265,990	88.4 / 11.6 / 0.0
TweetSentBR	14,784	7	7	5	1	51	14,400	44.6 / 29.8 / 25.6

AUC of 1, while a random model would have an AUC of 0.5.

4.4 Results and discussion

Tables 2 and 3 show the classification results from the models based on BERTimbau and twitter-XLM-roBERTa, respectively, for each dataset. The datasets are labeled with 2 classes (positive and negative), except for TweetSentBR, which also includes the neutral class. Therefore, with this dataset, we evaluated models with 3 and 2 classes (excluding neutral samples). The ROC-AUC metric was generated only for binary classification problems.

It can be seen that the classification based on the BERTimbau model performed better than the twitter-XLM-RoBERTa model in all datasets. This is due to the fact that the BERTimbau model was trained with data from the same domain from which it was evaluated, while the twitter-XLM-RoBERTa model is already pre-trained, and did not undergo fine-tuning on the datasets used.

Figure 1 shows the comparative graphic between the values obtained by the ROC-AUC metric for opinion mining approaches, for each dataset from [7]'s work. The two best results from the aforementioned work were selected (FastText 300 and Glove 300), of which the ROC-AUC metric was the only one used. The largest values in each dataset are highlighted in bold.

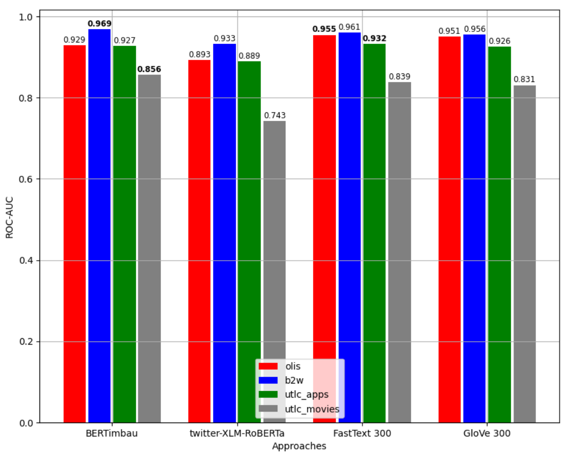

Figure 1: Results of opinion mining approaches for the ROC-AUC metric on datasets from [7]

The results show that our classifier based on the BERTimbau model obtained higher values in two of the datasets, while the LightGBM classifier using FastText embeddings of size 300, from the baseline [7], achieved higher values in the other two datasets. The twitter-XLM-RoBERTa model classifier obtained the lowest values in all cases, probably because it was trained with a tweet dataset, which has a different text style than the datasets used.

Figures 2 and 3 show the comparative graphics between the values obtained by the F1 and Accuracy metric for the opinion mining, for the TweetSentBR dataset from the work of [8] for 2 and 3 classes, respectively. It is worth mentioning that in the classification with 3 classes, the Hybrid Classifier approach was not used in the work of [8], and this work also did not use the ROC-AUC metric.

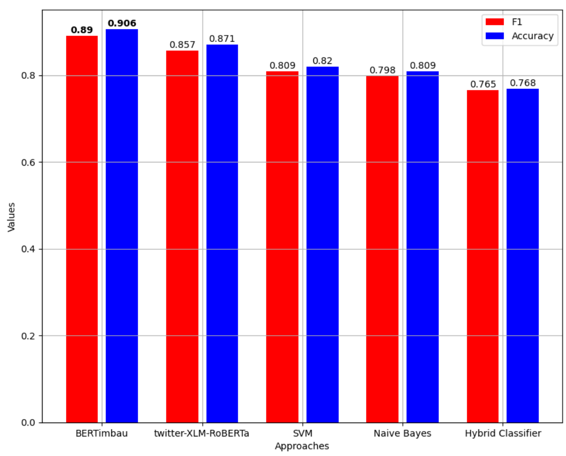

Figure 2: Results of opinion mining approaches for the F1 and Accuracy metrics on dataset TweetSentBR from [8] for 2 classes

The results show that our classifiers based on the BERT model obtained higher values than those from the work of [8] (SVM, Naive Bayes and Hybrid Classifier), for both 2 and 3 classes.

Given the results obtained, we can conclude that approaches based on the BERT model can be good options for the opinion mining task. Therefore, we will evaluate them in datasets about the cachaça beverage, as described in the next sections.

Table 2: Results for the models generated by BERTimbau fine-tuning

Dataset	Accuracy	Precision	Recall	F1	ROC-AUC
Olist	0.944	0.954	0.967	0.960	0.929
B2W	0.970	0.984	0.972	0.978	0.969
UTLC-Apps	0.946	0.968	0.962	0.965	0.927
UTLC-Movies	0.952	0.965	0.981	0.973	0.856
TweetSentBR - 2 classes	0.890	0.936	0.878	0.906	0.893
TweetSentBR - 3 classes	0.747	0.749	0.747	0.748	-

Table 3: Results for the twitter-XLM-roBERTa model

Dataset	Accuracy	Precision	Recall	F1	ROC-AUC
Olist	0.884	0.961	0.869	0.913	0.893
B2W	0.927	0.975	0.918	0.946	0.933
UTLC-Apps	0.861	0.979	0.838	0.903	0.889
UTLC-Movies	0.740	0.957	0.740	0.834	0.743
TweetSentBR - 2 classes	0.857	0.954	0.802	0.871	0.871
TweetSentBR - 3 classes	0.717	0.745	0.717	0.722	-

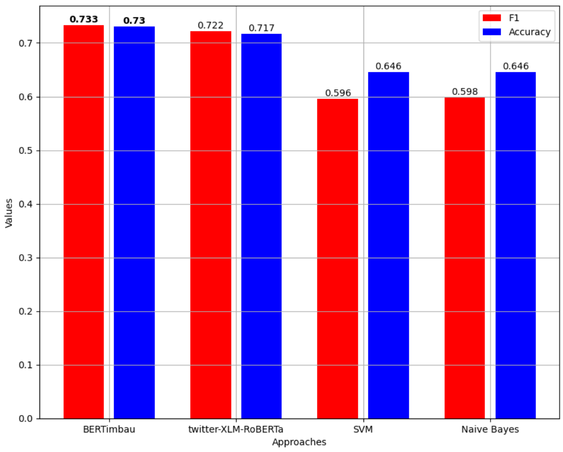

Figure 3: Results of opinion mining approaches for the F1 and Accuracy metrics on dataset TweetSentBR from [8] for 3 classes

5 Data Collection and Labeling of CachacaOM

In this section, we describe the process for collecting and labeling two datasets containing comments about the cachaça beverage, obtained from the social networks Facebook and X (formerly Twitter), for the opinion mining task. We call these datasets “CachacaOM” (Cachaça Opinion Mining) and describe the details of each collection and labeling in the following subsections.

5.1 Data Collection

To collect data from the social network Facebook, we used the GraphAPI[6] API, which is a tool made available by Facebook itself and allows for the developer to remove or insert information on the social network through an HTTPS protocol.

We extracted 34,620 comments, from 22,064 posts, from 19 pages related to the cachaça drink, referring to the sales sites used to extract information for the CachacaNER project [26]. The extraction occurred on the following pages: casadabebida, mbcachacaria, amburanabr, Sanhacu, lojacachacaepinga, cachacavelhobarreirooficial, cachacaepresente, domtapparoengenho, cachacarianacional, CachacaCompanheira, blubeer.com.br, araraunacachacaria, brme.oficial, cachacariasalinas.com.br, magnificadefaria, cachacasapucaia, wibacachaca, bebidaonline and cachacasbrasileirasoficial. From each page, all comments found were extracted, in all posts, from the creation of the page until September 2021. For each comment, the following were extracted: comment text, publication date, identifier and link.

The data was pre-processed, removing white spaces, tabs and line breaks, and excluding samples without text. Repeated samples were also disregarded. After preprocessing, the resulting dataset for Facebook has 34,440 samples.

To collect data from social network X, we used Twitter-API[7], which is a tool made available by X itself and allows us to remove or insert information on the social network via the HTTPS protocol.

We extracted 17,766 tweets from 11 pages related to the cachaça beverage and tweets that have the hashtag “#cachaca” in their text. The extraction took place on the following pages: araraunacachaca, BebaCompanheira, cachacanacional, cachacariasp, cachacaslinas, CachacaSapucaia, cachacawiba, DomTapparo, EmporioCCEma, NaBebidaOnline and sanhacu. For each tweet, the following were extracted: tweet text, publication date and identifier.

[6] https://developers.facebook.com/docs/graph-api

[7] https://developer.twitter.com/en/docs/twitter-api

After preprocessing, the resulting dataset for X has 17,765 samples.

5.2 Data Labeling

Data labeling is the process of assigning labels to raw data, which can be categorical, numerical or textual in nature. Labeling can be done manually or automatically. In this work, manual labeling was performed, using the doccano[8] tool, which is an open source tool for labeling text documents or images.

Due to the high cost of manual labeling, we only labeled a sample of the collected dataset. The labeled sample was randomly chosen from all collected instances. The labeling process was carried out by one of the authors of this work and supervised by another author. Each instance was labeled into one of the following classes: positive, negative, or neutral. A positive label is assigned to texts that express praise, positive reviews and desire for the product, for example: *"I also want that very much"* (In Portuguese, *"Eu também quero muito"*). Negative texts express dissatisfaction, negative reviews and complaints, for example: *"I'm waiting for it to arrive... I bought it on the 13th and they still haven't sent it to the carrier..."* (*"Tô esperando chegar... comprei dia 13 e ainda não enviaram pra transportadora..."*). Finally, the neutral label is for texts that express both positive and negative sentiment, advertisements, texts with only tags from other users of the platform, text that could not be categorized as positive or negative, for example: *"What is the price of this cachaça and how do I buy it?"* (*"Qual o preco dessa cachaças e como eu faço pra comprar"*).

We randomly selected 1,200 comments for labeling from the Facebook samples and 1,000 from the X samples. The result is presented in Table 4. Facebook data has a vocabulary of 2,570 words, with an average of 5 words per sample, median equal to 3, mode 2, minimum of 1 and maximum of 68 words. The X data has a vocabulary of 6,648 words, with an average of 17 words per sample, median equal to 16, mode 19, minimum of 1 and maximum of 53 words.

6 Evaluation of the CachacaOM Dataset

The second objective of this work was to evaluate the BERT model for the task of opinion mining in data about the cachaça beverage. As our labeled dataset CachacaOM is small for training and testing the models, it was used only for testing. The classifiers were trained using the datasets described in Section 4.1, obtained from the works of [7] and [8]. Thus, we are evaluating how good data from different domains are for training models for prediction in the cachaça beverage domain. The evaluation was done separately for data collected from Facebook and X.

Table 5 shows the results of the prediction of samples from the CachacaOM dataset, for the classes: positive and negative. The classifiers based on the BERTimbau model were generated by fine-tuning this model on each of the datasets, and the last line refers to the pre-trained model twitter-XLM-RoBERTa. For the TweetSentBR dataset, samples with a neutral label were excluded from training. The other datasets have only two classes. For the twitter-XLM-RoBERTa model, predictions for the neutral class were not accounted for. The values highlighted in bold represent the highest values for each of the Accuracy, Precision, Recall and F1 metrics for each of the models.

The results show that classifiers based on the BERTimbau model obtain higher values for the Accuracy and F1 metrics in relation to the twitter-XLM-RoBERTa model on Facebook data. The same does not happen for the X data, where the twitter-XLM-RoBERTa model obtains higher values in some cases. This is due to the fact that the training and testing data are of similar types, obtained from social network X, although they are from different domains. This also justifies the fact that the BERTimbau - TweetSentBR classifier obtains the highest values in the X data.

Table 6 shows the results of the prediction of samples from the CachacaOM dataset, for the classes: positive, negative and neutral. For this classification, only the BERTimbau - TweetSentBR and twitter-XLM-RoBERTa models were evaluated, as they are the only ones trained to predict three labels.

The classification into 3 classes proves to be a more challenging scenario. Both models in Table 6 were trained with data from X. For the cachaça data extracted from Facebook, the classifier based on the BERTimbau model obtained the highest values, while for the data extracted of X, the twitter-XLM-RoBERTa model obtained the highest values, probably due to the fact that the latter was trained with a larger volume of data. In any case, compared to the results presented in the last line of Tables 2 and 3, the CachacaOM dataset is more difficult to classify than the TweetSentBR dataset for 3 classes.

It is important to highlight that our dataset is small (2,200 instances) and very unbalanced, with a very low number of negative samples. This may limit our conclusions. However, this dataset was not used to train the models; it was used only for testing. The quantity is not sufficient to train and test models. The BERTimbau-based models were trained with data from Table 1. These data are also imbalanced, but in a much larger volume, with many instances of the negative class. They reflect the reality of opinion data, with around 70% positive and 30% negative opinions. Therefore, the trained models can be considered adequate for evaluation. We did a cross-domain evaluation.

We carried out an analysis of model error cases and present below some examples that demonstrate the dif-

[8] `https://github.com/doccano/doccano`

Table 4: Labels for Facebook and X data

	Facebook		X (Twitter)	
Label	#samples	Porcentage (%)	#samples	Porcentage (%)
Positive	576	48.0	103	10.3
Negative	31	2.58	19	1.9
Neutral	593	49.42	878	87.8

Table 5: Results for the dataset CachacaOM - 2 classes

	Facebook				X (Twitter)			
Model - Dataset	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
BERTimbau - Olist	0.945	0.977	0.965	0.971	0.901	0.909	0.980	0.943
BERTimbau - B2W	0.944	0.978	0.961	0.970	0.877	0.885	0.980	0.930
BERTimbau - UTLC-Apps	**0.947**	0.970	0.974	**0.972**	0.868	0.865	**1.000**	0.927
BERTimbau - UTLC-Movies	0.945	0.956	**0.987**	0.971	0.827	0.841	0.980	0.905
BERTimbau - TweetSentBR	0.929	0.987	0.937	0.961	**0.918**	0.926	0.980	**0.952**
twitter-XLM-RoBERTa	0.866	**0.997**	0.861	0.924	0.877	**0.978**	0.873	0.923

Table 6: Results for the dataset CachacaOM - 3 classes

	Facebook				X (Twitter)			
Model - Dataset	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
BERTimbau - TweetSentBR	**0.784**	**0.626**	0.730	**0.648**	0.755	0.487	0.638	0.522
twitter-XLM-RoBERTa	0.750	0.614	**0.767**	0.629	**0.780**	**0.498**	**0.742**	**0.542**

ficulty of classifying the CachacaOM dataset. For the data extracted from Facebook, the comment "Great! However, I'm still waiting for mine for August. It's taking a while, CN!..." ("Excelente! Porém ainda estou à espera da minha do mês de Agosto. Está demorando, CN!...") is a negative sample. However, almost all classifiers incorrectly predicted this instance, probably due to the use of the more positive term "Excellent". Another example is the comment "not available for purchase yet??" ("não está disponivel pra compra ainda??"), which is a positive example, given the interest in purchasing the product. However, almost all classifiers mislabeled this instance as negative, possibly due to the use of the negation term "not".

For the X data, we found that all classifiers erroneously labeled as positive the sample "#CACHAÇA and #PERFUME are champions in #TAXES. Cachaça, I can't confirm. But #PERFUME, it's really quite salty. But it's worth it, it's #fragrant" ("#CACHAÇA e #PERFUME sao campeãs,em #IMPOSTOS. Cachaça,n posso confirmar.Mas #PERFUME,realmente tá bem salgado.Mas vale a pena,ser #cheirosa"). However, it is a negative sample, as it is a complaint. Another example is "I tried to translate "love of my life", the answer was CACHAÇA!" ("Fui tentar traduzir "amor da minha vida", a resposta foi CACHAÇA!"), which is a positive sample, given the customer's appreciation for the drink. However, it has been wrongly labeled by some models as negative.

To conclude, according to our experimental evaluation, the BERT model has the potential to be used in opinion mining in data about the cachaça beverage. Furthermore, in the absence of data from the cachaça domain for training the models, the datasets used in this work demonstrate good results between different domains. As a result of the evaluation, for classification into 2 classes, it is recommended to use the BERTimbau - UTLC-Apps model for data extracted from Facebook and the BERTimbau - TweetSentBR model for data from X. For classification into 3 classes, it is recommended model BERTimbau - TweetSentBR for Facebook data and the twitter-XLM-RoBERTa model for X data.

7 Conclusion

In this study, we evaluate approaches based on the BERT model for opinion mining in Portuguese. We also collected and labeled a dataset with comments about the cachaça beverage. Furthermore, we carried out an experimental evaluation of classification models trained in different domains to predict data about the cachaça beverage.

In evaluating the BERT model for opinion mining in the Portuguese language, we fine-tuned the BERTimbau model on several datasets collected from e-commerce platforms and X (Twitter), from different domains. We also use a pre-trained model on X data for opinion mining. The results demonstrate that approaches based on the BERT model obtained higher values in the ROC-AUC, Accuracy and F1 metrics in most evaluations when compared to two baselines, with values greater than 0.97 in the F1 metric for classification into 2 classes, and 0.74 for 3 classes.

We collected a dataset containing 52,205 comments about the cachaça beverage, extracted from the social networks Facebook and X. We labeled a random sample of 2,200 comments from this dataset for opinion mining, in the positive, negative and neutral classes. Although the labeled dataset is small, it represents an important contribution, as no labeled data on the cachaça beverage was found in the literature for the opinion mining task.

In the cross-domain evaluation, models trained in different domains were used for evaluation on our labeled dataset about cachaça. The results demonstrate that approaches based on the BERT model have potential for opinion mining in the cachaça beverage domain, with values greater than 0.97 being achieved in the F1 metric for classification into 2 classes and 0.64 for 3 classes. They are, therefore, an alternative when you do not have large labeled datasets in the cachaça domain.

As future work, we intend to expand the labeling of the cachaça dataset. There are still many samples that can be labeled in our data collected from social media. It is important that we have more labeled data, especially in the negative class, to make it sufficient for training machine learning models. We also intend to evaluate other machine learning models for opinion mining about cachaça, especially the large open source language models focused on the Portuguese language. LLMs have proven effective in sentiment analysis. In their research, [41] show that ChatGPT excels at determining opinion polarity and can be used to label datasets. In [42], the authors investigated ChatGPT's ability to automate lexicon annotation.

Authors' contribution

Thiago Santos implemented the codes, labeled the dataset and carried out the experiments. Mozar Brito supervised the content related to cachaça and wrote an excerpt of the text. Denilson Pereira supervised the work, reviewed the dataset and wrote the text of the article.

Competing interests

The authors have declared that no competing interests exist.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript.

Data Availability

The repository with the code and data is publicly available.[9]

Funding

This work was partially supported by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq) and the Minas Gerais Research Support Foundation (Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG).

Acknowledgements

We would like to thank Arthur Silveira Franco and Priscilla de Souza Silva for his help with data collection.

References

[1] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, 2022. [Online]. Available: https://doi.org/10.1007/s10462-022-10144-1

[2] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys*, vol. 50, no. 2, pp. 25:1–25:33, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3057270

[3] D. A. Pereira, "A survey of sentiment analysis in the portuguese language," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 1087–1115, 2021. [Online]. Available: https://doi.org/10.1007/s10462-020-09870-1

[4] W. Zhang, Y. Deng, B. Liu, S. J. Pan, and L. Bing, "Sentiment analysis in the era of large language models: A reality check," *arXiv preprint arXiv:2305.15005*, 2023.

[5] Instituto Brasileiro da Cachaça, "IBRAC," 2022, accessed in September, 2025. [Online]. Available: https://ibrac.net/

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*. Minneapolis, Minnesota: Association for Computational Linguistics, jun 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[7] F. D. Souza and J. B. de Oliveira e Souza Filho, "Sentiment analysis on brazilian portuguese user reviews," *arXiv preprint arXiv:2112.05459*, 2021.

[8] H. Brum and M. d. G. Volpe Nunes, "Building a sentiment corpus of tweets in Brazilian Portuguese," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), May 2018, pp. 4167–4172. [Online]. Available: https://aclanthology.org/L18-1658

[9] L. da Câmara Cascudo, *Prelúdio da Cachaça*. São Paulo: Global Editora, 2015.

[10] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013. [Online]. Available: https://doi.org/10.1109/MIS.2013.30

[11] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, p. 102–107, mar 2016. [Online]. Available: https://doi.org/10.1109/MIS.2016.31

[9] https://github.com/ThiagoSallesSantos/IC_AnaliseSentimento/

[12] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Information Fusion*, vol. 44, pp. 65–77, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566253517303901

[13] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artificial Intelligence Review*, vol. 56, pp. 8469—-8510, August 2023. [Online]. Available: https://doi.org/10.1007/s10462-022-10386-z

[14] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014. [Online]. Available: https://doi.org/10.1016/j.asej.2014.04.011

[15] J. C. F. Neto, D. A. Pereira, B. H. G. Barbosa, and D. D. Ferreira, "Approaches based on language models for aspect extraction for sentiment analysis in the portuguese language," *Neural Computing and Applications*, vol. 36, pp. 19 353–19 363, November 2024. [Online]. Available: https://doi.org/10.1007/s00521-024-10265-4

[16] L. Zhang and B. Liu, *Aspect and entity extraction for opinion mining*. Springer, 2014, book Chapter, Volume 1 of the series Studies in Big Data.

[17] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: tasks, methods, and challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11 019–11 038, november 2022. [Online]. Available: https://doi.org/10.1109/TKDE.2022.3230975

[18] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence-level sentiment analysis," *Information Sciences*, vol. 512, pp. 1078 – 1102, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025519309879

[19] B. Cardoso and D. A. Pereira, "Evaluating an aspect extraction method for opinion mining in the portuguese language," in *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*. Porto Alegre, RS, Brasil: SBC, 2020, pp. 137–144. [Online]. Available: https://sol.sbc.org.br/index.php/kdmile/article/view/11969

[20] A. A. L. Cunha, M. C. Costa, and M. A. C. Pacheco, "Sentiment analysis of youtube video comments using deep neural networks," in *International Conference on Artificial Intelligence and Soft Computing (ICAISC)*. Cham: Springer International Publishing, 2019, pp. 561–570. [Online]. Available: https://doi.org/10.1007/978-3-030-20912-4_51

[21] R. P. da Silva, F. A. O. Santos, F. B. do Nascimento, and H. T. Macedo, "Cross-language approach for sentiment classification in brazilian portuguese with convnets," in *Information Technology - New Generations*, S. Latifi, Ed. Cham: Springer International Publishing, 2018, pp. 311–316.

[22] V. T. F. Kuwaki, M. N. Ladeira, M. G. G. Benitez, and R. J. T. Junior, "Building a corpus from supermarket reviews in portuguese for document-level sentiment analysis," *Anais do XIII Computer on the Beach*, vol. 13, pp. 119–125, May 2022.

[23] M. Won and J. Fernandes, "Ss-pt: A stance and sentiment data set from portuguese quoted tweets," *Lecture Notes in Computer Science*, vol. 13208, pp. 110–121, March 2022. [Online]. Available: https://doi.org/10.1007/978-3-030-98305-5_11

[24] F. D. Souza and J. B. d. O. e. S. Filho, "Bert for sentiment analysis: Pre-trained and fine-tuned alternatives," in *International Conference on Computational Processing of the Portuguese Language*, vol. 13208. Springer, 2022, pp. 209–218.

[25] F. Souza, R. Nogueira, and R. Lotufo, "Bertimbau: Pretrained bert models for brazilian portuguese," in *Intelligent Systems*, R. Cerri and R. C. Prati, Eds. Cham: Springer International Publishing, 2020, pp. 403–417. [Online]. Available: https://doi.org/10.1007/978-3-030-61377-8_28

[26] P. Silva, A. Franco, T. Santos, M. Brito, and D. A. Pereira, "Cachacaner: a dataset for named entity recognition in texts about the cachaça beverage," *Language Resources and Evaluation*, 2023, online Publication. [Online]. Available: https://doi.org/10.1007/s10579-023-09665-0

[27] D. Calbino, M. J. de Brito, and V. d. G. P. Brito, "Reordenação do status da cachaça de alambique: uma abordagem sob a ótica do trabalho institucional," *Revista Eletrônica de Ciência Administrativa*, vol. 21, no. 1, pp. 37–66, 2022. [Online]. Available: https://doi.org/10.21529/RECADM.2022002

[28] J. E. de Souza, E. R. Scharf, and G. A. Gehrke, "Identidade de marca de cachaças artesanais: Um gole pro santo!" *Revista Interdisciplinar de Marketing*, vol. 12, no. 1, pp. 52–68, 2022. [Online]. Available: https://periodicos.uem.br/ojs/index.php/rimar/article/view/61185

[29] E. T. T. de Araújo, J. K. L. Silva, F. C. P. dos Santos, and A. C. Ferreira, "O consumo de cachaça e seus sentidos: uma análise do comportamento do consumidor à luz da teoria do sensemaking," *Revista Gestão Organizacional*, vol. 14, no. 2, pp. 46–68, 2021. [Online]. Available: https://doi.org/10.22277/rgo.v14i2.5392

[30] B. U. Rodrigues, R. M. Costa, R. L. Salvini, A. A. Soares, F. A. Silva, M. Caliari, K. C. R. Cardoso, and T. I. Ribeiro, "Cachaça type identification using color information and computer vision," in *X Workshop de Visão Computacional*, vol. 10, 2014, pp. 45–49.

[31] G. C. Silvello, A. M. Bortoletto, M. C. de Castro, and A. R. Alcarde, "New approach for barrel-aged distillates classification based on maturation level and machine learning: A study of cachaça," *LWT*, vol. 140, p. 110836, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0023643820318259

[32] S. S. Virnodkar, V. K. Pachghare, V. C. Patil, and S. K. Jha, "Application of machine learning on remote sensing data for sugarcane crop classification: A review," in *ICT Analysis and Applications*,

S. Fong, N. Dey, and A. Joshi, Eds., vol. 93. Singapore: Springer Singapore, 2020, pp. 539–555. [Online]. Available: https://doi.org/10.1007/978-981-15-0630-7_55

[33] "Brazilian e-commerce public dataset by olist," 2018, https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce. Accessed in November, 2023.

[34] L. Real, M. Oshiro, and A. Mafra, "B2W-Reviews01 - an open product reviews corpus," in *Proceedings of the XII Symposium in Information and Human Language Technology*, Salvador, BA, October 2019, pp. 200–208. [Online]. Available: https://github.com/b2wdigital/b2w-reviews01

[35] N. S. Hartmann, L. V. Avanço, P. P. Balage, M. S. Duran, M. d. G. V. Nunes, T. A. S. Pardo, and S. M. Aluisio, "A large corpus of product reviews in portuguese: tackling out-of-vocabulary words," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*. Reykjavik, Iceland: European Language Resources Association (ELRA), May 2014, pp. 3865–3871. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/413_Paper.pdf

[36] R. F. d. Sousa, H. B. Brum, and M. d. G. V. Nunes, "A bunch of helpfulness and sentiment corpora in brazilian portuguese," in *Symposium in Information and Human Language Technology - STIL*, Salvador, BA, October 2019, pp. 209–218.

[37] J. A. Wagner Filho, R. Wilkens, M. Idiart, and A. Villavicencio, "The brWaC corpus: A new open resource for Brazilian Portuguese," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki, Japan: European Language Resources Association, May 2018. [Online]. Available: https://aclanthology.org/L18-1686

[38] F. Barbieri, L. Espinosa Anke, and J. Camacho-Collados, "XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, Jun 2022, pp. 258–266. [Online]. Available: https://aclanthology.org/2022.lrec-1.27

[39] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul 2020, pp. 8440–8451. [Online]. Available: https://aclanthology.org/2020.acl-main.747

[40] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized bert pretraining approach," *ArXiv*, vol. abs/1907.11692, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:198953378

[41] G. de Araujo, T. de Melo, and C. M. S. Figueiredo, "Is chatgpt an effective solver of sentiment analysis tasks in portuguese? a preliminary study," in *Proceedings of the 16th International Conference on Computational Processing of Portuguese-Vol. 1*, 2024, pp. 13–21.

[42] F. S. Marcondes, A. d. C. O. S. Gala, M. Rodrigues, J. J. Almeida, and P. Novais, "Lexicon annotation with llm: A proof of concept with chatgpt," in *Hybrid Artificial Intelligent Systems*, H. Quintián, E. Corchado, A. Troncoso Lora, H. Pérez García, E. Jove Pérez, J. L. Calvo Rolle, F. J. Martínez de Pisón, P. García Bringas, F. Martínez Álvarez, Á. Herrero, and P. Fosci, Eds. Cham: Springer Nature Switzerland, 2025, pp. 190–200.

Citation: T. S. Santos, M. J. Brito and D. A. Pereira. *Evaluation of Approaches Based on the BERT Model for Opinion Mining about the Cachaça Beverage*. Journal of Computer Science & Technology, vol. 25, no. 2, pp. 107-117, 2025.
DOI: 10.24215/16666038.25.e09.
Received: July 8, 2025 **Accepted:** October 10, 2025.
Copyright: This article is distributed under the terms of the Creative Commons License CC-BY-NC-SA.