Procesamiento del Lenguaje Natural, Revista nº 66, marzo de 2021, pp. 111-120 recibido 30-11-2020 revisado 24-01-2021 aceptado 04-02-2021

Consumer Cynicism Identification for Spanish Reviews using a Spanish Transformer Model

Identificación del cinismo del consumidor para reseñas en español utilizando un modelo de transformador español

Samuel González-López[1], Steven Bethard[2], Francisca Cecilia Encinas Orozco[3], Adrián Pastor López-Monroy[4]
[1]Technological University of Nogales
[2]University of Arizona
[3]University of Sonora
[4]Mathematics Research Center (CIMAT)
sgonzalez@utnogales.edu.mx, bethard@arizona.edu,
cecilia.encinasorozco@unison.mx, pastor.lopez@cimat.mx

Abstract: Companies pay close attention to how consumers react on social media to their products or services. Our work focuses on the identification of Consumer Cynicism, defined as a negative attitude that can have a broad or specific focus and comprises cognitive, affective, and behavioral components. We create a corpus of 619 Spanish-language comments on YouTube car reviews, annotated for four cynicism constructs: Dissatisfaction, Alienation, Skepticism, and Hostility. We compare different classification formulations (binary vs. multi-label) and different pre-trained models (Spanish BETO vs. multilingual BERT). We find binary classifiers derived from BETO consistently outperform multi-label classifiers and classifiers derived from BERT. Our best models achieve F1 of 0.83 for Dissatisfaction, 0.77 for Hostility, 0.71 for Skepticism and 0.70 for Alienation.
Keywords: Consumer Cynicism, binary classification model, multi-label model, social media.

Resumen: Las empresas prestan mucha atención a las reacciones de los consumidores de sus productos o servicios en las redes sociales. Nuestro trabajo se centra en la identificación del cinismo del consumidor, el cual se define como una actitud negativa que puede tener un enfoque amplio o específico y comprende los componentes cognitivo, afectivo y conductual. Creamos un corpus de 619 comentarios en el idioma español sobre reseñas de automóviles de YouTube, los comentarios se etiquetaron para cuatro constructos del cinismo: Insatisfacción, Alienación, Escepticismo y Hostilidad. Además, comparamos diferentes formulaciones de clasificación (binaria vs. multi-etiqueta) y diferentes modelos pre-entrenados (BETO-español vs. BERT-multilingüe). Encontramos que los clasificadores binarios derivados de BETO superan consistentemente a los clasificadores de etiquetas múltiples y a los clasificadores derivados de BERT. Nuestros mejores modelos alcanzan F1 de 0.83 para Insatisfacción, 0.77 para Hostilidad, 0.71 para Escepticismo y 0.70 para Alienación.
Palabras clave: Cinismo del Consumidor, modelo de clasificacion binaria, modelo multi-etiqueta, redes sociales.

1 Introduction

The need to predict customers' behavior has led companies to carry out studies on opinions in digital media. Brands seek to position themselves and provide satisfaction to their potential customers. The global adspend growth for this year is 20% reaching an investment of $84 billion in social media (Santini et al., 2020). An analysis developed from 48 brands in 8 industries (including car brands) showed that volume metrics explain brand awareness and purchase intent. The volume

ISSN 1989-7553. DOI 10.26342/2021-66-9

Components	Related Constructs
Cognitive	Suspicion Mistrust Skepticism Distrust
Affective	Alienation Dissatisfaction
Behavioral	Resistance Hostility

Table 1: Components of consumer cynicism.

metrics correspond to a collection of the number of likes, comments, and shares of posts on Facebook and YouGov (Kübler, Colicev, and Pauwels, 2020).

Many opinions users write on different social media platforms can be classified as negative, neutral, or positive. Works related to sentiment analysis in comments have received significant interest (Kauffmann et al., 2019). In some studies deep learning techniques have been used to classify reviews (Tammina and Annareddy, 2020; Kocoń, Miłkowski, and Zaśko-Zielińska, 2019). However, at a deeper level of analysis, we can find other behaviors that can differ from each other despite being negative or positive.

Our research seeks to identify behaviors related to consumer cynicism. Cynicism is defined as a negative attitude that can have a broad or specific focus and comprises cognitive, affective, and behavioral components (Chylinski and Chu, 2010). Table 1 shows the components and constructs of consumer cynicism defined by prior work (Chylinski and Chu, 2010). Our work focuses on analyzing the following subset of those constructs:

- Skepticism: Doubt of consumer related of the product or brand.
- Alienation: Consumer feels disillusioned, powerless, hopeless, detached from the product or brand.
- Hostility: Consumer attempts to force alternative or desired features on the product.
- Dissatisfaction: Unmet expectations by the consumer, negative perception of the product or brand.

We explore different configurations of the Bidirectional Encoder Representations from Transformers (BERT; (Devlin et al., 2019)) models to classify car reviews on YouTube in the Spanish language. Our contributions are the following:

- We annotate a corpus of YouTube car reviews in Spanish considering the four selected cynicism constructs.
- We train BERT models for Skepticism, Alienation, Dissatisfaction, and Hostility, reaching 0.71, 0.70, 0.83, and 0.77 F1, respectively.
- We demonstrate that binary classification models outperform multi-label models for these problems.

2 Related work

Work closely related to the study of cynicism has analyzed behaviors such as offensive language, sarcasm, irony, and aggression.

Within offensive language, we find the analysis of profanity, insults, and abuse, and explicit or implicit offensive language in German-language tweets using BERT (Risch et al., 2019). This work reports an F1 of 51.2 % for three subcategories of offensive language and 73.1 % for explicit/implicit language. Techniques such as NaiveBayes and Support Vector Machines have also been explored, obtaining accuracy results of 92 % and 90 % (De Souza and Da Costa-Abreu, 2020), but the analysis only sought to identify the offensive language category. A comparison between the Perspective tool (a Convolutional Neural network CNN) and BERT found that on the SEMEVAL2019-Offenseval dataset, the Perspective tool had better performance identifying the offensive language category, while BERT had better performance identifying the insult, threat, and attack offensive language elements (Nikolov and Radivchev, 2019). Similarly, in SEMEVAL2020-Offenseval, the use of BERT + CNN in the analysis of offensive language in social networks found results above those of traditional techniques (Safaya, Abdullatif, and Yuret, 2020).

Recent work has also analyzed the concept of sarcasm. In an analysis of 21 papers on sarcasm, 22.58 % of the cases used Support Vector Machine (SVM) as an analysis technique, followed by 19.35 with Logistic Regression, 9.67 % Naive Bayes (Sarsam et al., 2020). Recently other works have explored BERT models to identify sarcasm, first extracting local features of words in sentences and later implementing a CNN to summarize all sentences (Srivastava et al., 2020).

Closely related to Sarcasm, we find Irony analyzed under approaches such as CNN with Embeddings (FastText, Word2vec) (Ghanem et al., 2020). This work analyzed monolingual and multilingual architectures in three languages, obtaining better performance from the monolingual configuration. Another approach, RCNN-RoBERTa, consists of a RoBERTa pre-trained transformer followed by a bidirectional Long short-term memory (BiLSTM), reaching 0.80 F1 in the SemEval-2018 dataset and 0.78 F1 in the Reddit Politics dataset (Potamias, Siolas, and Stafylopatis, 2020).

Related work has also attempted the identification of aggression. Aggression can be direct or indirect and is a feeling of anger that results in hostile behavior. Under the BERT framework and an assembly strategy, a dataset labeled as non-aggressive, covertly aggressive, and overtly aggressive was classified, and the assemblies achieved two percentage points higher F1-score than single models (Risch and Krestel, 2020). Using the same dataset but with other training features, for example, the amount of abusive/aggressive/offensive words or the presence of hash-tags, these features were incorporated into a CNN to obtain an accuracy of 73.2 % (Kumar et al., 2020).

Though we take inspiration from this prior work, our work differs in task and language. We focus on cynicism, annotating a new corpus for several previously unexplored constructs. And in contrast to most previous work that focused on the English language, our consumer cynicism analysis is on the Spanish language, and we consider both monolingual Spanish and multilingual configurations of BERT models.

3 Dataset

The training and test data set focus on user comments on YouTube channels with car review topics in the Spanish language. Reviews were pulled from 26 videos from five different car review channels. The review comments were filtered considering two requirements:

1. Comment size: comments should contain at least ten words. The goal here is to ensure sufficient text to judge the presence or absence of the four constructs of the current study. This parameter was set after the annotators performed qualitative analysis on the comments.
2. Relevance: comments should have a minimum of 5 likes. The goal here is to focus on comments that a substantial number of users find interesting.

Category	Measure	Score
Dissatisfaction	Cohen	0.88
Alienation	Cohen	0.79
Skepticism	Cohen	0.81
Hostility	Cohen	0.82
Overall	Fuzzy	0.79

Table 2: Two-annotators Kappa Agreement.

The resulting comments were tagged by two annotators who had knowledge related to consumer marketing. Each annotator was instructed to annotate each comment with one or more of Dissatisfaction, Alienation, Skepticism, Hostility, or to annotate the comment as None if none of these were present. Each annotator received a guide that defined each construct and provided annotated examples. These guidelines[1] can be found in appendix A. FindingFive[2] was used for the annotation process, and annotators' responses were automatically collected.

The filters reduced the number of comments to be tagged by annotators. For example, one of the videos with the highest number of comments exceeded 1250 but the filters reduced this to 90. After applying the filters to all comments from all 26 videos, 725 comments remained.

Table 2 shows annotator agreement on the 725 comments, measured with Kappa's Cohen (Landis and Koch, 1977) and Fuzzy Kappa (Dou et al., 2007). The fuzzy measure produces a single value for all categories, and allows comments to belong to more than one category. All measures of agreement reached 0.79 or higher, indicating good agreement.

We retained for training and evaluation data the 619 out of 725 filtered comments where annotators agreed on the labels. Table 3 shows the distribution of the categories

[1] The guide is displayed in the original language of the study. Suitable for Spanish annotators.

[2] `https://www.findingfive.com/`

Category	Percentage
Dissatisfaction	16
Alienation	20
Skepticism	18
Hostility	17
None	29

Table 3: Label distribution in the annotated corpus.

in this subset, and Table 4 shows some annotated examples from the corpus.

After the tagged process, we found comments where the annotators had disagreements. For instance, comments were tagged with the category none by the first annotator, while the second annotator selected dissatisfaction or alienation construct. Below are a couple of examples.

1. *Mi Subaru Legacy 1993 1ra Generación es mas espacioso, mas equipado y mas cómodo*/'Mi Subaru Legacy 1993 1ra Generación es mas espacioso, mas equipado y mas cómodo'. Tagged as None/Dissatisfaction.

2. *Lo bueno del march es que es el más económico de todos y en equipamiento está 'completo', digo por el precio que tiene el march*/'The good thing about the march is that it is the cheapest of all and in terms of equipment it is 'complete', I mean for the price of the march'. Tagged as None/Alienation.

These types of examples were not used in the experimentation.

4 Methodology

The collected corpus was used to train machine-learning models. We compare two different formulations of the classification problem: training one binary classification model for each construct of cynicism (Dissatisfaction, Alienation, Skepticism, and Hostility) vs. training a single multi-label model that predicts all constructs jointly. We also compare two different pre-trained transformer models[3] from which cynicism models can be fine-tuned: the multilingual BERT model (Devlin et al., 2019), and the Spanish-language BETO model (Cañete et al., 2020). We also consider adding a convolutional neural network (CNN) layer before the output of the above models. The rest of this section describes these options in detail.

BERT The Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2019) uses a transformer neural network to infer contextual representations of the words in a text. BERT uses a masked language modeling objective to pre-train the transformer network on large unlabeled data: the BooksCorpus (800M words), and English Wikipedia (2,500M words)

Multilingual BERT The BERT training paradigm has been used to train a variety of models that vary in their sizes and their training data. We select for our purposes bert-base-multilingual-cased, a model with 12 self-attention layers, 12 attention heads, 768-dimensional word representations, and 110M parameters. It was trained on cased text in the top 104 languages with the largest Wikipedias.

Spanish BERT (BETO) The BETO model (Cañete et al., 2020) is trained using the BERT training paradigm, and is similar in size to bert-base-multilingual-cased. However, BETO focuses only on the Spanish language. We use BETO-cased that has 12 self-attention layers, 16 attention heads each, 1024-dimensional word representations, and 110M parameters. It was trained on Spanish Wikipedia and the sources of the OPUS Project that had text in Spanish. These sources include the United Nations and Government journals, TED talks, subtitles, and News Stories. According to the authors of BETO, the total of corpora used is comparable to the original BERT. The comparison of BETO vs. BERT multilingual for Spanish gave favorable results to BETO (Cañete et al., 2020).

Convolutional network Though it is possible to make predictions directly from a BERT-style model, we also consider taking the contextual word representations from BERT or BETO and feeding them to a convolutional neural network (CNN) to make the predictions.

Data processing The corpus comments used were delimited by the punctuation mark

[3]In preliminary experiments we also explored bag of words and lemmatization features with naive Bayes, logistic regression, and LSA algorithms, but the BERT-based models always outperformed these.

Spanish Example	English Translation	Labels
se ve buena la suv, poco a poco parecen convencer más que otras marcas por el costo beneficio en cuestión de equipamiento, el único problema sería la confiabilidad que otras marcas como honda dan pero hasta no ver no creer.	'the SUV looks good, but slowly they seem to convince more than other brands for the cost-benefit in equipment issue, the only problem would be the reliability that other brands like honda give but until you see, not believe.'	Skepticism
Es culpa de Nissan y también de la gente, lo siguen comprando por ser la opción fácil y "segura". Mientras los consumidores no exijan mejores productos seguirán existiendo autos como estos.	'It is the fault of Nissan and the people. They continue to buy it for being the easy and safe option. As long as consumers do not demand better products, cars like these will continue to exist.'	Hostility
Nissan es una de las compañías automotriz más grandes en el mundo y aún no pueden introducir una nueva generación en todos los mercados.	'Nissan is one of the world's largest automotive companies, and they cannot yet introduce a new generation in all markets.'	Dissatisfaction Skepticism

Table 4: Label examples from the annotated corpus.

point, which indicated that the comment was finalized. The rest of the symbols and words were kept throughout the entire process. The elements of the test group were randomly selected.

5 *Results and Analysis*

The collected corpus was divided into three partitions: training, validation, and testing. The first dataset was used to train each of the implemented models, one binary and multilabel. The validation set was used to fine-tune the model hyperparameters. We use the validation set results to perform the analysis for section five. Finally the test set was used to run the best configuration of the BERT model. The hyper-parameters that were adjusted during the evaluation of the validation set were the batch size, the number of training epochs and the learning rate starting with a value of 5e-5.

5.1 Dissatisfaction construct

The results of experimentation for the Dissatisfaction construct are detailed in Table 5. Multi-label models with convolutional layers (rows 3 and 4) outperformed multi-label models without (rows 1 and 2). All binary classification models (rows 5 and 6) outperformed all multi-label models (first 4 rows). And for the better models (last 4 rows), the Spanish BETO models (rows 4 and 6) outperformed their corresponding multilingual BERT models (rows 3 and 5, respectively). The best configuration was Spanish BETO + binary classification, obtaining an F-measure of 0.83.

When manually reviewing the false positives for the dissatisfaction construct, we found local words or words typical to a geographic region that affected the prediction.

> *La Ram esta muy genial pero esa suspensión de aire no va a aguantar la friega...*
>
> 'The Ram is very great, but the air suspension will not withstand hard work...'

or

> *Jeep no está diseñado para chulearlo en la plaza...*
>
> 'Jeep is not designed to be displayed in the square...'

We can see in the first example the *friega* word that means hard work and represents wear for the suspension. In the second example, the *chulearlo* word appears, which means provide positive words about some object or thing.

5.2 Alienation construct

The Alienation construct results in Table 6 showed similar trends as dissatisfaction. Convolutional layers improved the performance of multi-label models, all binary classification models outperform all multi-label models, and for the better models (last 4 rows), Spanish BETO models outperformed multilingual BERT models. The best configuration was

Language	Model	P	R	F
Multi	Multi	0.18	0.52	0.27
Spanish	Multi	0.19	0.51	0.28
Multi	Multi+CNN	0.38	0.48	0.42
Spanish	Multi+CNN	0.63	0.50	0.56
Multi	Binary	0.76	0.63	0.69
Spanish	Binary	0.92	0.78	**0.83**

Table 5: Dissatisfaction construct results in terms of Precision (P), Recall (R), and F. Language is either Multilingual BERT or Spanish BETO. Model is either individual Binary models or a Multi-label model, and may include a convolutional (CNN) layer before the output.

Language	Model	P	R	F
Multi	Multi	0.34	0.51	0.41
Spanish	Multi	0.27	0.45	0.34
Multi	Multi+CNN	0.44	0.53	0.48
Spanish	Multi+CNN	0.45	0.75	0.56
Multi	Binary	0.52	0.72	0.61
Spanish	Binary	0.64	0.76	**0.70**

Table 6: Alienation construct results in terms of Precision (P), Recall (R), and F. Language is either Multilingual BERT or Spanish BETO. Model is either individual Binary models or a Multi-label model, and may include a convolutional (CNN) layer before the output.

again Spanish BETO + binary classification, obtaining an F-measure of 0.70.

Manual analysis of the development set reveals the presence of short expressions of popular wisdom annotated as Alienation. For instance:

A cualquier santo moderno le rezan...

'They pray to any modern saint...'

or

El que no conocio a Dios en el pasado...

'He who did not know God in the past...'

This construct obtained the lowest level of Kappa agreement. Incorporating a method for the detection of theses phrases in the early stages of training could help improve results.

5.3 Hostility construct

Like the preceding results, Table 7 shows that for the hostility construct binary classification outperforms multi-label models and the Spanish BETO models outperform the multilingual BERT model. However, unlike the preceding results, adding a convolutional layer to a multi-label model does not consistently yield an improvement. The best configuration was again Spanish BETO + binary classification, obtaining an F-measure of 0.77.

Language	Model	P	R	F
Multi	Multi	0.35	0.46	0.40
Spanish	Multi	0.52	0.59	0.55
Multi	Multi+CNN	0.42	0.55	0.48
Spanish	Multi+CNN	0.39	0.75	0.51
Multi	Binary	0.68	0.73	0.70
Spanish	Binary	0.76	0.79	**0.77**

Table 7: Hostility construct results in terms of Precision (P), Recall (R), and F. Language is either Multilingual BERT or Spanish BETO. Model is either individual Binary models or a Multi-label model, and may include a convolutional (CNN) layer before the output.

When manually reviewing the data, we found comments with dissatisfaction content that includes negative phrases in a pejorative word game close to sarcasm. For example:

La suspensión trasera la cagaron, mejor una suspensión trasera independiente como las generaciones anteriores. Pero los Mazdetos felices con cualquier cosa

'The rear suspension was shit, better an independent rear suspension like previous generations. But the Mazdetos are happy with anything'

If the experts' labeling of the Hostility construct includes the sarcasm subcategory, then we could identify with some level this type of word game. Negative comments focused on the presenter were not taken into account in this construct. The performance of BETO on this construct was only below the performance of the Dissatisfaction construct. It would appear that BERT models work best with negative comments.

5.4 Skepticism construct

Table 8 shows that, similar to hostility, for the skepticism construct binary classification outperforms multi-label models, the Spanish BETO models outperform the multilingual BERT model, and there is no consistent benefit to adding a convolutional layer to multi-label models. The best configuration was

Language	Model	P	R	F
Multi	Multi	0.26	0.48	0.34
Spanish	Multi	0.32	0.48	0.38
Multi	Multi+CNN	0.31	0.43	0.36
Spanish	Multi+CNN	0.53	0.64	0.58
Multi	Binary	0.56	0.77	0.65
Spanish	Binary	0.64	0.80	**0.71**

Table 8: Skepticism construct results in terms of Precision (P), Recall (R), and F. Language is either Multilingual BERT or Spanish BETO. Model is either individual Binary models or a Multi-label model, and may include a convolutional (CNN) layer before the output.

again Spanish BETO + binary classification, obtaining an F-measure of 0.71

When reviewing the false positives, we found comments with the behaviors described above in the Alienation and Hostility constructs. For instance:

> *Este auto me recuerda a nissan sentra mas bien puro humo.*
>
> 'This car reminds me of a Nissan Sentra, rather pure smoke'

We also found that many comments have a particular negative load, similar to comments of dissatisfaction, for example:

> *Ya sabemos como sera de mediocre y feo el auto con solo ver el emblema de la marca*
>
> 'We already know how mediocre and ugly the car will be just by looking at the brand's emblem'

The dissatisfaction comments specify a dislike for the price or some component of the car, while for the alienation construct, the comments are more global and show a particular dislike of the brand or the car, such as the car's reliability.

6 Discussion

The overall results are summarized in Table 9. Models for all four constructs achieve an F1 of at least 0.70, with the dissatisfaction model achieving the highest F1 of 0.83. Across all constructs, we observed similar patterns: binary classification outperformed multi-label classification and the Spanish BETO model outperformed the multilingual BERT model.

In addition to the metrics reported in the results section, we review the values produced by the best configuration for all constructs: Spanish BETO + binary classification. For each construct, we gather the models' predictions over all test items and plot the distribution using a box plot. Figure 1 shows the result. An ideal result would be that all models have probabilities far from 0.5, suggesting high confidence in both the presence and absence of a construct. The Alienation model has the largest range of probabilities, while the Skepticism model has the smallest range. In fact, the Skepticism model's probabilities are all very close to 0.5, suggesting that the model is rarely confident whether a comment represents skepticism or not. Levels of agreement on labeling may need to improve. The kappa value for the Dissatisfaction construct was the highest, and in Figure 1, Dissatisfaction obtains the highest probability values.

Category	F-measure
Dissatisfaction	0.83
Hostility	0.77
Skepticism	0.71
Alienation	0.70

Table 9: Constructs F-measure results.

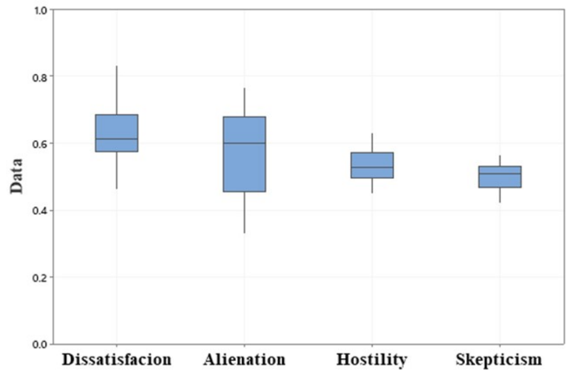

Figure 1: Test groups probabilities.

7 Conclusion

The use of pre-trained models is beneficial since they significantly cover relationships between words due to their enormous amounts of training data. In this study, we have analyzed different configurations of BERT pre-trained models. We found the best performance from binary classification models based on BETO. We also found that for all four constructs we studied, the Spanish language BETO model outperformed the multilingual BERT model.

Of the consumer cynicism constructs, Dissatisfaction obtained the best performance (0.83 F1) and Alienation the lowest (0.70 F1). We believe these F1 scores are encouraging, given the modest amount of data we annotated and the fact that we do not perform any additional pre-processing for our models.

Our analysis suggested that detecting sarcasm and identifying typical terms for geographic regions may be important aspects of future work on cynicism models. It may also be worth exploring hierarchical models that first detect the cynicism component (e.g., Affective) before the cynicism construct (e.g., Alienation). Expanding the corpus, both in terms of size, and in terms of variety of components covered, is also an important direction for future work. Overall, we believe that consumer cynicism is a little-explored concept, which could be of great interest to brands, due to the marked trend of markets towards social media platforms.

References

Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.

Chylinski, M. and A. Chu. 2010. Consumer cynicism: antecedents and consequences. *European Journal of Marketing*, 44(6):796–837.

De Souza, G. A. and M. Da Costa-Abreu. 2020. Automatic offensive language detection from twitter data using machine learning and feature selection of metadata. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6.

Dehghani, M., M. K. Niaki, I. Ramezani, and R. Sali. 2016. Evaluating the influence of youtube advertising for attraction of young customers. *Computers in Human Behavior*, 59:165–172.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Dou, W., Y. Ren, Q. Wu, S. Ruan, Y. Chen, D. Bloyet, and J.-M. Constans. 2007. Fuzzy kappa for the agreement measure of fuzzy classifications. *Neurocomputing*, 70(4):726 – 734. Advanced Neurocomputing Theory and Methodology.

Ghanem, B., J. Karoui, F. Benamara, P. Rosso, and V. Moriceau. 2020. Irony detection in a multilingual context. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, editors, *Advances in Information Retrieval*, pages 141–149, Cham. Springer International Publishing.

Kauffmann, E., J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora. 2019. Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining. *Journal of Sustainability*, 11:1 – 19.

Kocoń, J., P. Miłkowski, and M. Zaśko-Zielińska. 2019. Multi-level sentiment analysis of PolEmo 2.0: Extended corpus of multi-domain consumer reviews. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 980–991, Hong Kong, China, November. Association for Computational Linguistics.

Kumar, R., A. K. Ojha, S. Malmasi, and M. Zampieri. 2020. Evaluating aggression identification in social media. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 1–5, Marseille, France, May. European Language Resources Association (ELRA).

Kübler, R. V., A. Colicev, and K. H. Pauwels. 2020. Social media's impact on the consumer mindset: When to use which sentiment extraction tool? *Journal of Interactive Marketing*, 50:136 – 155.

Landis, J. R. and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Nikolov, A. and V. Radivchev. 2019. Nikolov-radivchev at SemEval-2019 task 6: Offensive tweet classification with BERT and ensembles. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 691–695, Minneapolis, Minnesota, USA, June. Association for Computational Linguistics.

Potamias, R., G. Siolas, and A. Stafylopatis. 2020. A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, pages 1433 – 3058.

Risch, J. and R. Krestel. 2020. Bagging BERT models for robust aggression identification. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 55–61, Marseille, France, May. European Language Resources Association (ELRA).

Risch, J., A. Stoll, M. Ziegele, and R. Krestel. 2019. hpidedis at germeval 2019: Offensive language identification using a german bert model. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*, pages 405–410, Erlangen, Germany. German Society for Computational Linguistics & Language Technology.

Safaya, A., M. Abdullatif, and D. Yuret. 2020. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media.

Santini, F. D. O., W. Ladeira, D. Pinto, M. Herter, C. Sampaio, and B. Babin. 2020. Customer engagement in social media: a framework and meta-analysis. *Journal of the Academy of Marketing Science*, 48(6):1211–1228.

Sarsam, S. M., H. Al-Samarraie, A. I. Alzahrani, and B. Wright. 2020. Sarcasm detection using machine learning algorithms in twitter: A systematic review. *International Journal of Market Research*, 62(5):578–598.

Srivastava, H., V. Varshney, S. Kumari, and S. Srivastava. 2020. A novel hierarchical BERT architecture for sarcasm detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 93–97, Online, July. Association for Computational Linguistics.

Tammina, S. and S. Annareddy. 2020. Sentiment analysis on customer reviews using convolutional neural network. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6.

A Appendix: Annotation Guidelines in Spanish

Guía de anotación para identificar "cinismo del consumidor" en reviews de YouTube.

Introducción: En la actualidad las redes sociales juegan un papel relevante para las empresas que ofrecen productos o servicios. Un estudio realizado a 378 estudiantes de la universidad de Sapienza Roma, se enfocó al análisis de elementos que pudieran contribuir a la intención de compra en videos de YouTube. Los usuarios promediaban de 1 a 7 horas de videos y el 98% tenía experiencia en redes sociales. El análisis reportó que la personalización y el entretenimiento percibido en los videos eran relevantes para la intención de compra (Dehghani et al., 2016). Los videos bajo esta plataforma permiten la interacción con los usuarios, a través de comentarios u opiniones de estos. En específico nuestra investigación busca identificar comportamientos relacionados al cinismo del consumidor, éste se define como una actitud negativa que puede tener un enfoque amplio y especifico, y se conforma de los componentes cognitivo, afectivo y comportamental (Chylinski and Chu, 2010). A continuación, se muestran los tres componentes y sus constructos en la Table 10.

Componentes	Constructos relacionados
Cognitivo	Ecepticismo
Afectivo	Alienación Insatisfacción
Comportamental	Hostilidad

Table 10: Componentes del cinismo del consumidor.

De la Table 10 el constructo "Hostilidad" se refiere cuando un cliente/usuario busca resultados alternos o deseados. Este estudio se enfoca en las opiniones y/o comentarios escritos por usuarios en canales de reviews de autos en español. Por ejemplo, en el siguiente comentario de un usuario de YouTube, se puede apreciar el deseo de que se realice una crítica del auto más a fondo: "Hola Gabo espero que cuando la camioneta llegue a México la critiques como debe ser, se entiende que los chinos te invitaron y no se puede decir más...un abrazo de chile".

Instrucciones del etiquetado: El objetivo del etiquetado es identificar si un comentario se relaciona a los siguientes constructos Escepticismo, Alienación, Insatisfacción u Hostilidad. A continuación se muestran ejemplos de cada constructo.

Ejemplo del constructo Escepticismo: Se ve buena la suv, poco a poco parecen convencer más que otras marcas por el costo beneficio en cuestión de equipamiento, el único problema sería la confiabilidad que otras marcas como honda dan pero hasta no ver no creer.

Ejemplo del constructo Alienación: Es culpa de Nissan y también de la gente, lo siguen comprando por ser la opción fácil y "segura". Mientras los consumidores no exijan mejores productos seguirán existiendo autos como estos.

Ejemplo del constructo Insatisfacción: Yo tengo un jetta, pero por 422,900 jajajajajajajaja mejor le completo y me compro un bmw XD esta muy caro el jetta.

Ejemplo de constructo Hostilidad: que equivocado está señor yo tengo una Cadillac y créame que es muy superior a Mercedes y a BMW su motor y el lujo es muy superior y es más grande que sus rivales.

A continuación se muestra un ejemplo etiquetado.

Comentario	Constructo
se ve buena la suv, poco a poco parecen convencer más que otras marcas por el costo beneficio en cuestión de equipamiento, el único problema sería la confiabilidad que otras marcas como honda dan pero hasta no ver no creer.	*Escepticismo* Alienación Insatisfacción Hostilidad
yo tengo un jetta, pero por 422,900 jajajajajajajaja mejor le completo y me compro un bmw XD esta muy caro el jetta.	Escepticismo Alienación *Insatisfacción* Hostilidad

Table 11: Ejemplo anotado.

Como se aprecia en la Table 11, el primer comentario fue etiquetado como Escepticismo, mientras que el segundo comentario se le asignó la etiqueta de Insatisfacción. El proceso de anotación se realizará a través de la plataforma FindingFive. El acceso a la plataforma se enviará por correo electrónico.