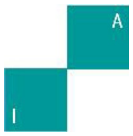
Inteligencia Artificial 28(76) (2025), 301-310
doi: 10.4114/intartif.vol28iss76pp301-310



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Leveraging Transformers Ensemble to Improve Aspect Mining in Spanish Reviews

Miguel Ángel Rivero Tapia[1], Alfredo Simón-Cuevas[1,A] and Orlando Grabiel Toledano-López[2]

[1]Universidad Tecnológica de La Habana José Antonio Echeverría, CUJAE, Ave. 114, e/Rotonda y Ciclovía, La Habana, Cuba
[A]asimon@ceis.cujae.edu.cu
[2]Universidad de las Ciencias Informáticas (UCI), Carretera a San Antonio de los Baños Km 2½, Rpto. Torrens, La Habana, Cuba

Abstract Aspect-based sentiment analysis is a process aimed to understanding the sentiment expressed in opinions or reviews about specific features of an entity. The automatic extraction of aspects is the most challenging task, as it requires the ability to understand the context and to recognize the relevant and characteristic elements of an entity about which you have an opinion. To increase the quality results in the solution to this problem is still a challenge in Spanish reviews, because very few papers have been reported and the reported efficacy rates need to be improved. The use of deep learning models has proven an advantage for aspect extraction task, but the combination of several models for obtaining a final prediction has not yet been exploited. In this work, an aspect extraction method in which several Transformer models are combined through an ensemble learning approach using the Average Voting technique is presented. The proposed solution was evaluated using the SemEval2016 dataset and the results obtained were compared to those reported by other state-of-the-art solutions. The evaluation process not only provides a starting point to have a broader perception of the performance of the Transformers in this context, but also highlights the improvement of the quality results of the aspect extraction with the Transformers-Based Ensemble.

Keywords: Aspect-based sentiment analysis, aspect extraction, Transformers models, ensemble learning.

1 Introduction

The large volume of textual information in user-generated opinions or reviews on the web (e.g., social networks, blogs, chats, news, and e-commerce platforms, etc.) has become a key focus of attention for business, governmental, and social organizations. The automatic processing of this information can reveal how users and citizens think about the products and services provided, public policies, and certain events or actions. These criteria may prove valuable knowledge for decision-making in several scenarios.

Sentiment analysis (or opinion mining, as it is also known) is the field aimed at the computational study of people's opinions, feelings, ratings, attitudes, and emotions, towards entities such as products, services, organizations, individuals, problems, events, issues and their characteristics or attributes [1]. Opinions refer to a person's point of view towards a certain issue. In contrast, the sentiment or polarity refers to the emotion experienced by the person towards that issue in the text, which is usually identified as positive or negative. Aspect-based sentiment mining (ABSA) is one of the categories or lines of work within sentiment analysis, where computational processing is taken to the lowest level of granularity, where entities, aspects, and relationships present in opinions are identified and classified separately [2]. Monitoring and analyzing the reviews of users, customers, or citizens in general with this approach is essential from a business and governmental perspective, in the current scenario where the content on social networks has such an impact on the decisions of citizens.

ABSA solutions are divided into two fundamental tasks: aspect extraction and aspect-based sentiment classification. The first part aims to identify and extract the aspect about which an opinion is being issued. The

ISSN: 1137-3601 (print), 1988-3064 (on-line)

second part is responsible for detecting the sentiment polarity associated with that aspect of the opinion [3]. The automatic aspects extraction is the most challenging task, as it requires the ability to understand the context and recognize the relevant and characteristic elements of an entity about which an opinion or judgment is expressed. The interest in ABSA solutions has grown in recent years due to the need to understand user feedback more deeply. However, developments and evaluations have mainly been addressed in English reviews. The aspects extraction from Spanish reviews have been less attention to; consequently, there are few reported solutions and Spanish datasets available [4]. Supervised learning approaches are the most widely used to deal with the automatic aspect extraction challenges, and the deep learning models are the ones that have gained the most recognition [5], [6]. However, this trend is not reflected in the solutions reported for automatic aspect extraction from Spanish reviews, although some reported works using such approaches [7]–[9], the application of Transformer models is limited [8] and there are no reported solutions that apply the integration of Transformer for aspect extraction. Ensemble learning is a promising alternative for Transformers integration because it combines robustness to noise in the data with computational efficiency comparable to individual models [10]. This work aims to extend the experimental study of the application of Transformer models and their integration through ensemble architectures, for identifying a solution that improves the effectiveness of aspect extraction in Spanish-language reviews.

This paper presents a new aspect extraction method in which Transformer-based ensemble learning is proposed for extracting aspects from Spanish reviews. In this solution, three Transformer models (BETO + BERT + ALBERT-large) are integrated into an ensemble architecture, and the Average Voting technique is used for combining the predictions of the models and obtaining a final prediction. The proposed method was reached after studying and evaluating 10 individual Transformer models, as well as several integration variants and ensemble techniques. The experimental evaluation was carried out using the SemEval2016 reviews dataset [11], which is highly regarded in this task, and one of the few existing Spanish datasets [4]. The results obtained were compared with those reported by other state-of-the-art approaches, reflecting the contribution of the proposed method to increase the effectiveness of the aspect extraction from Spanish texts. Furthermore, the experiment conducted can also be useful as a baseline for performance evaluation of Transformer-based automatics aspect extraction.

The remainder of the paper is structured as follows: Section 2 presents a description and synthesis of the relevant aspects of the related works; Section 3 describes the proposed solution; Section 4 presents and analyses the results of the experimental evaluation process; and the conclusions reached and elements to be addressed in the future are presented in Section 5.

2 Related Works

ABSA plays a crucial role in applications such as customer feedback analysis, market research, and product and service improvement. The new potential and challenges offered by the ABSA solutions have captured the increasing attention of researchers worldwide [5]. This kind of solution stands as a sophisticated tool in the field of sentiment analysis, offering a deeper and more granular understanding of the opinion expressed in a text. In contrast to the simple classification of positive or negative sentiment, the ABSA focuses on analyzing the polarity of the sentiment toward specific characteristics or aspects, of a product or service, and certain events or actions. In this context, an ‘aspect’ is defined as a relevant attribute that contributes to the evaluation object.

ABSA solutions are divided into two fundamental tasks: aspect extraction and aspect-based sentiment classification [3], [6]. The first is to identify the specific characteristics or attributes, called "aspects" that are mentioned in a text. For example, in a restaurant review, the aspects could be the quality of the food, the service, the decoration, and the price, among others. The second is that, once the aspects have been identified, the polarity of the sentiment expressed toward each of them must be determined. In other words, the objective is to determine whether the opinion on the aspect is positive, negative, or neutral. The automatic extraction of aspects is the most challenging task, as it requires the ability to understand the context and identify the relevant and characteristic elements on which an opinion of an entity is expressed. This research focuses precisely on this problem.

The application of ABSA solutions has boomed in recent years, driven by the need to understand users' views in greater depth, leading to the development of a broad diversity of problem-solving approaches. Supervised approaches are the most widely used in the solution of this problem. These include traditional methods such as Support Vector Machine (SVM) and Naive Bayes, which are used for both polarity detection and aspect extraction, as well as approaches based on deep neural networks and transformer models, which have become more popular in recent years [5], [6]. Although several solutions reported in the literature, there are still many shortcomings in this field, including the absence of solutions targeted and evaluated in languages other than English. Most of the existing works are designed and evaluated for the English language, so there is an absence of solutions reported and applied to texts in other languages such as Spanish [7], [12], [13]. In this sense, it is of great

relevance to carry out studies for the conception of solutions and their evaluation in the field of reviews in Spanish; a solution with good results in English would not necessarily also achieve good results in Spanish. Resolving these linguistic limitations is crucial to expanding the scope and effectiveness of aspect-based sentiment analysis in various linguistic contexts. On the other hand, existing systems are not completely effective in extracting aspects of Spanish-language opinions.

In [12] an unsupervised ontology-based model for explicit and implicit aspect detection is proposed, and a similar approach is also used to classify aspect polarity. The unsupervised approach facilitates the scalability of the system to other languages and domains because it does not require training processes, but the quality of the results is not high, since the reported F1 results are 0,73. In [8] the use of a pre-trained language model is proposed, specifically the BETO model (a model based on BERT, but pre-trained specifically for Spanish language tasks), and reported 0,767 of F1-score on the Semeval 2016 dataset. In [7] also considered the use of deep learning models, specifically combining two Convolutional Neural Networks (CNNs) architectures for aspect detection. However, one of the problems with the CNNs is that they do not take into account the order of the words in the sentence, so they are not fully effective in capturing the meaning of the words in context [14]. The results they reported using SemEval 2016 are 0,654 as an F1-score.

3 Aspect Extraction Method

The proposed method aims to extract the most relevant aspects from Spanish reviews. The method is conceived of two main phases: (1) Preprocessing and (2) Aspect Extraction, where the latter is carried out by applying a Transformers-based ensemble, as shown in Fig. 1. The combination of Transformers and the ensemble technique adopted was the one that obtained the best results after experimental evaluations with different variants.

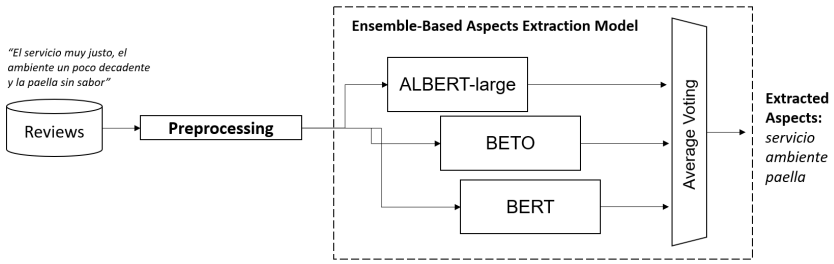


Figure 1. Pipeline of the aspect extraction method.

1.1 Preprocessing

Aspect-based sentiment analysis typically requires customized preprocessing to prepare the text data and extract information effectively. However, preprocessing steps for Transformer-based models differ significantly from traditional approaches. While techniques such as converting text to lowercase, removing punctuation, or suppressing stop words are common for models like SVMs (Support Vector Machine) or LSTMs (Long Short-Term Memory), they are often unnecessary or even detrimental for Transformers. This is because these models are pre-trained on raw text that includes such elements, and their tokenizers are designed to handle them natively.

The first step in our preprocessing pipeline was the removal of special characters and encoding normalization to ensure text compatibility. Nevertheless, the most critical and substantial preprocessing step was transforming the raw text into the specific input format required by Transformer models. This involved using each model's respective tokenizer to convert sentences into sequences of subword tokens, along with the addition of special tokens [CLS] and [SEP] as required by the models.

For the model training phase, it was necessary to align the tokenized sequences with the ground truth labels. Since these models are commonly used for token classification, we formulated the task as a sequence labeling problem. An example of this representation is shown in Fig. 2. For instance, in the review "The food was tasty," the relevant aspect is "food." After tokenization, a label of 1 was assigned to each token that constituted an aspect, while a label of 0 was assigned to all other tokens. This created the necessary aligned binary representations for supervised fine-tuning.

```
text_tokens;tags;polarities
['La', 'comida', 'estuvo', 'sabrosa', '.'];[0, 1, 0, 0, 0];[-1, 1, -1, -1, -1, -1]
['La', 'Carne', ',', 'fenomenal'];[0, 1, 0, 0];[-1, 1, -1, -1]
['Mal', 'Servicio'];[0, 1];[-1, 0]
```

Figure 2. Example of the result of preprocessing an opinion.

1.2 Aspect extraction

Aspect extraction was addressed by applying a Transformers-based ensemble model, which refers to a method in which multiple base models are integrated into the same framework to obtain a more robust classification model. The effectiveness of an ensemble method depends on several factors, including how the base models are trained and how their predictions are combined [15]. There are several techniques for predicting a result that can be applied in an ensemble approach, such as Voting, Bagging, Stacking, and Boosting. In this paper, we propose an approach based on the Average Voting Ensemble, integrating the BETO + BERT + ALBERT-large. The method takes advantage of the complementary strengths of each Transformer through a probabilistic consensus mechanism.

Initially, each Transformer model is tuned using the SemEval 2016 review corpus, adapting the parameters to the specific aspect extraction task. During this process, a final classification layer is set up to generate binary probabilities (0-1) corresponding to the Aspect and Non-Aspect labels. Subsequently, the calibrated models process the test set of the same corpus, producing for each token a vector of probabilities reflecting its association with the target class. In the final stage, the outputs of the three models for each token are aggregated by arithmetic averaging. This operation generates a consolidated vector where each value represents the consensus probability that a token constitutes an aspect. Fig. 3 illustrates an example of aspects extraction from the review: '*El servicio muy justo, el ambiente un poco decadente y la paella sin sabor*'. In this example, the average of predictions results in a high probability for the tokens *'servicio'*, *'ambiente'* and *'paella'*, so they are identified as aspects.

Transformers Models	*El*	*servicio*	*muy*	*justo*	*,*	*el*	*ambiente*	*un*	*poco*	*decadente*	*y*	*la*	*paella*	*sin*	*sabor*
Beto	00,07	99,83	00,08	00,07	00,02	00,47	99,75	00,16	00,20	00,04	00,02	00,12	99,53	00,30	00,11
Bert	03,12	99,32	00,38	00,12	00,15	00,13	97,33	00,37	00,16	00,18	00,28	01,12	97,48	00,32	09,30
Albert_Large	00,16	99,23	00,03	00,02	00,11	00,09	99,02	00,08	00,06	00,15	00,02	00,16	90,36	00,12	19,28
Average Voting Ensemble	01,11	99,46	00,16	00,07	00,09	00,23	98,70	00,20	00,14	00,12	00,10	00,46	95,79	00,24	09,56

Figure 3. Example of the aspect identification via Average Voting Ensemble of BETO + BERT + ALBERT-large.

The proposed ensemble method combines BETO, BERT and ALBERT-large through average voting to improve the robustness of aspect extraction by leveraging their complementary strengths as evaluated in the proposed solution: the Spanish specialization of BETO, the general comprehensiveness of BERT and the computational efficiency of ALBERT. This approach seeks to smooth outliers through probabilistic consensus and benefits from domain-specific fine tuning of each ensemble model. By averaging its predictions, the method attempts to mitigate individual errors to achieve greater accuracy and robustness in feature extraction than any single model. A probabilistic consensus can provide more reliable results, although it requires more computational resources. This strategy overcomes the limitations of monolithic approaches, especially in texts with specialized vocabulary, and seeks synergy between models.

4 Evaluation and Discussion

The experimental evaluation of the solution was carried out with the dataset of Spanish reviews offered at SemEval-2016 for Task 5: Aspect-Based Sentiment Analysis [11]. This collection is made up of restaurant reviews, labeled with the aspects that should be identified in an automatic process (for example: service, food, atmosphere, attention, among others). This dataset is the main and mostly used Spanish dataset for the evaluation of aspect extraction solutions, due to the lack of availability of resources in the form of datasets and tools in the Spanish language [4]. The results were computed using the Precision, Recall, and F1-score metrics. Table 1 shows a characterization of the ABSA-SemEval-2016 dataset.

Table 1: Characterization of the ABSA-SemEval-2016.

	Reviews	Test	Train	Aspects
ABSA-SemEval-2016	2697	627	2070	247

The experimental evaluation was carried out on a total of 10 pre-trained Transformers models: BERT multilingual-cased [16], BETO (BERT-based model, but pre-trained specifically for Spanish-language tasks), ELECTRA two versions (Spanish version) [17], BERTIN (and its variants) [18], GPT-2-Spanish [19], and AlBERT (in three different sizes) [20]. Unlike BETO, specifically for the Spanish language, BERT multilingual was prepared in different languages, including Spanish, Chinese, Arabic, Hindi, Portuguese, French, and others. A fine-tuning process was applied to each of the models, to adapt them as a solution to the problem of extracting aspects from reviews. Each model was trained during 5 epochs and a batch size or batch of 8. The choice of batch size was motivated by several factors: large batch sizes can lead to fluctuations in gradient updates. Larger batch sizes can lead to faster convergence, but there is a risk that the model will be overly tight to the training data and not generalize well into new data. Table 2 shows the configuration of the different hyperparameters defined for the execution of the experiments with each of the models.

A robust ensemble requires models that make different errors. To achieve this, it is essential to incorporate models that are diverse in their fundamental design. The present selection encompasses the primary sources of diversity among transformer models: Diversity based on language, Diversity based on architecture and pre-training objective, GPT-like (Decoder-only) models, and Diversity based on implementation. Furthermore, for an ensemble to be effective, the individual models must be sufficiently competent. The selected models include the most prominent and well-regarded architectures for Spanish language processing.

The selection of models was based on their diversity of architectures, training techniques, and optimizations, which allows for covering a broad spectrum of needs. BERT was chosen for laying the groundwork of contextual learning. ALBERT was included for its efficiency in reducing parameters without compromising performance, while RoBERTa demonstrates that more robust pre-training enhances the potential of the BERT architecture. GPT-2 contributes its unique ability to generate coherent and long-form text, and ELECTRA introduces a more efficient training method that outperforms BERT. Together, this structural and methodological variety ensures there is an optimized model for any practical scenario, whether prioritizing accuracy, speed, or resource efficiency.

The experimentation goals were as follows:

1. To evaluate each of the models independently, in order to have a perception of the quality of the results of each model in the extraction of aspects;
2. To evaluate different combinations of Transformers models under an ensemble learning approach.

Table 2: Hyperparameters configuration.

Hyperparameter	Value
Seed	50
Learning rate	$1e^{-5}$
Epochs	5
Batch size	8
Maximum word sequence size	512

The results from the first task are shown in Table 3, highlighting in bold the highest values of each evaluation metric. The ALBERT-large model obtains the highest value of F1. Therefore, ALBERT-large is the best overall performance of the evaluated models. However, in terms of Precision, the following models stand out with the highest values: BETO, BERT, and ALBERT-xx-large. Some models, such as BERTIN-large and ELECTRA-base show consistent results across all metrics, given that both have Precision and Recall around 84 % and 81 %, making them reliable, but they don't stand out as much as ALBERT's models. GPT-2 is the worst performer, particularly in Recall, suggesting that it struggles to correctly identify the true positives. Table 4 presents the results of evaluating three integration schemes of Transformers using five ensemble learning techniques: Stacking (STK), Boosting (BOOS), Maximum Voting (MXVT), Average Voting (AVT), and Weighted Average Voting (WAVT).

Table 3: Results of Transformers models in aspect extraction.

Id.	Models	Precision	Recall	F1
1	BETO	**86.36**	83.46	84.83
2	BERT	85.17	83.13	84.11
3	ALBERT-base	82.38	83.00	82.69
4	ALBERT-large	84.68	85.79	**85.22**
5	ALBERT-xx-large	82.23	**86.27**	84.14
6	BERTIN-base	78.31	83.12	80.44
7	BERTIN-large	84.03	81.45	82.67
8	GPT-2	78.24	53.39	52.82
9	ELECTRA-small	81.27	80.06	80.65
10	ELECTRA-base	82.85	82.29	82.57

Table 4: Results of different Transformer combination schemes and ensemble techniques.

Ensemble variants	Transformer Combination Schemes	Precision	Recall	F1
STK-1	BETO + BERT + ALBERT_Large	86,36	83,46	84,83
STK-2	BETO + ALBERT_Large + ALBERT_xx_Large	87,38	82,73	84,85
STK-3	BETO + BERT + BERTIN_Large + ALBERT_Large + ALBERT_xx_Large	85,37	85,26	85,31
BOOS-1	BETO + BERT + ALBERT_Large	88,38	82,80	85,30
BOOS-2	BETO + ALBERT_Large + ALBERT_xx_Large	85,69	85,71	85,66
BOOS-3	BETO + BERT + BERTIN_Large + ALBERT_Large + ALBERT_xx_Large	87,71	83,81	85,62
MXVT-1	BETO + BERT + ALBERT_Large	87,61	85,16	86,33
MXTV-2	BETO + ALBERT_Large + ALBERT_xx_Large	85,77	85,92	85,84
MXTV-3	BETO + BERT + BERTIN_Large + ALBERT_Large + ALBERT_xx_Large	87,09	85,84	86,45
AVT-1	BETO + BERT + ALBERT_Large	**90,63**	88,19	**89,36**
AVT-2	BETO + ALBERT_Large + ALBERT_xx_Large	88,67	88,91	88,79
AVT-3	BETO + BERT + BERTIN_Large + ALBERT_Large + ALBERT_xx_Large	89,82	88,69	89,24
WAVT-1	BETO + BERT + ALBERT_Large	88,35	88,69	88,52
WAVT-2	BETO + ALBERT_Large + ALBERT_xx_Large	88,61	**89,11**	88,86
WAVT-3	BETO + BERT + BERTIN_Large + ALBERT_Large + ALBERT_xx_Large	89,38	88,64	89,05

The Transformers selected were the ones with the highest F1 or Precision results in the individual evaluations, specifically the Top 5. Another element considered in the selection was the necessary diversity in the ensemble composition. The ensemble of BETO + BERT + ALBERT_large through Average Voting technique (AVT-1) is

the best performer overall, given that it reached the highest F1 (89,36 %) and Precision values (90,63 %); only variant with Precision higher than 90 %. This solution is the most effective in terms of the balance between Precision and Recall, and our proposal for improving the effectiveness of the aspect extraction task. The WAVT-2 variant reached the highest value of Recall (89,11 %), overcoming 9 % to AVT-1.

The results show that the application of a Transformer ensemble improves the effectiveness of the aspect extraction task, compared to the use of individual Transformer models. The overall F1 results of the different combinations are higher than the average F1 of the individual results of the Transformer models that comprise them. Specifically, the F1 obtained by AVT-1 is 5% higher than the average F1 of the individual Transformers results. According to Fig. 4 and 5, Average Voting is the most promising ensemble technique because it offers a better balance between Precision and Recall, and higher F1. On the other hand, Stacking and Boosting perform slightly less well but may be viable options depending on the context.

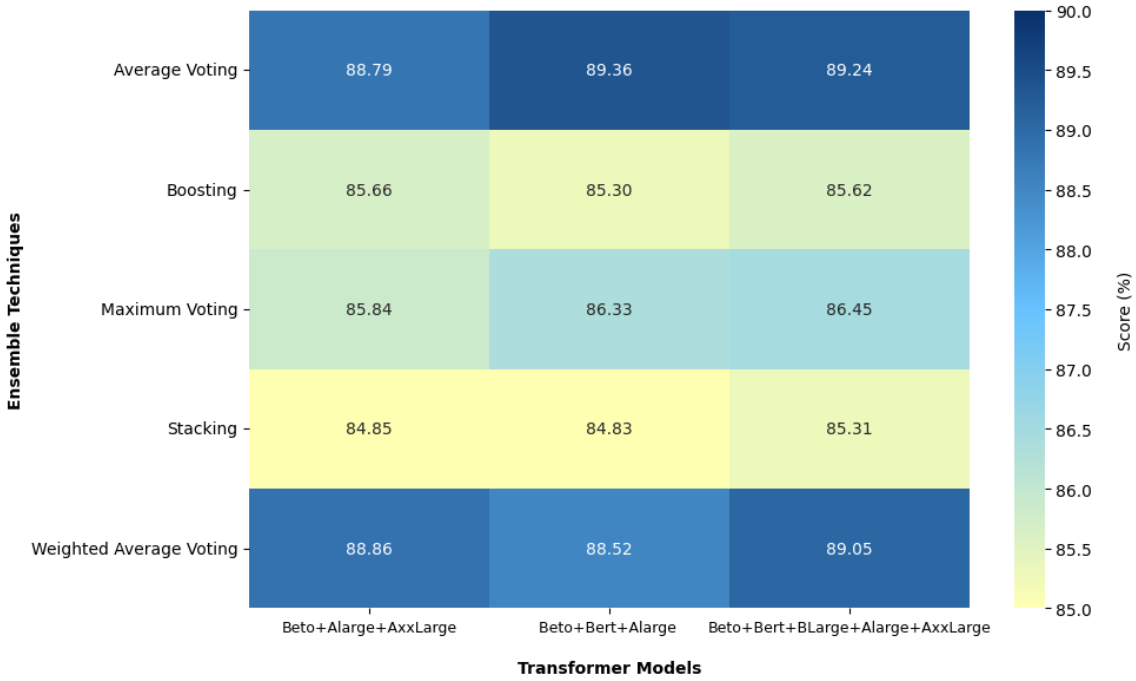


Figure 4. Performance of different ensemble techniques according F1 results.

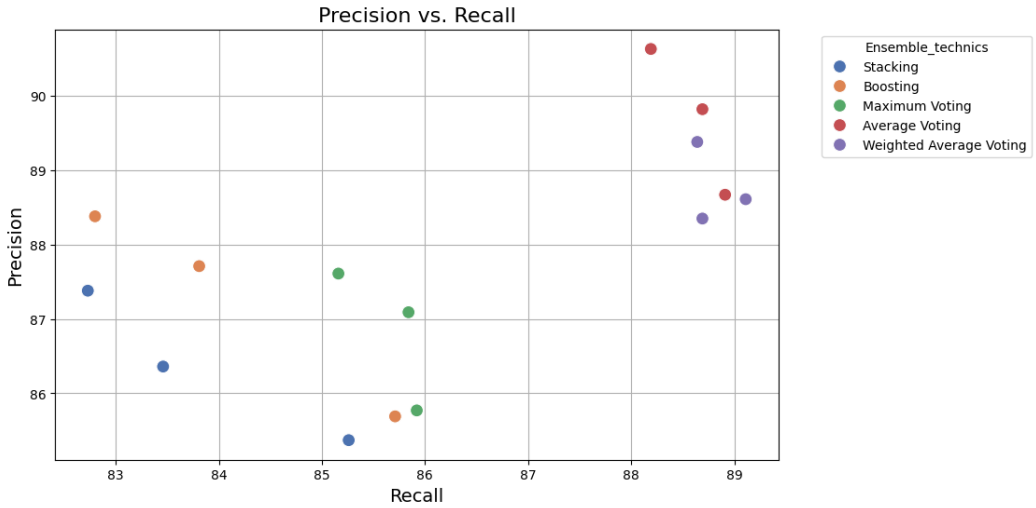


Figure 5. Balance between Precision and Recall for each ensemble technique.

Fig. 6 shows that all combinations evaluated have high performance, with average values of F1 and Precision above 86,80 % and 87,22 %, respectively. An important conclusion from this comparative analysis is that increasing the number of models to be integrated into an ensemble architecture does not necessarily produce a

representative improvement of the effectiveness. In this experiment the combination of Beto + Bert + Bertin_Large + Albert_Large + Albert_xx_Large has slightly better performance in F1 but may require more computational resources than Beto + Bert + Albert_Large, therefore, the last combination would be recommended.

Table 5 shows the comparison of the results of the proposed method with those obtained by other solutions reported and evaluated using the same dataset, using F1 results as a reference metric. Three of the compared proposals use a deep learning approach for aspect extraction, specifically, Convolutional Neural Network [7], BiLSTM-CRF [9], and BETO model [8]. In this comparison, it can be appreciated that the application of ensemble learning with Transformers outperforms the rest of the solutions reported with an appreciable improvement of 20 %, with respect to the average F1 values of all other solutions.

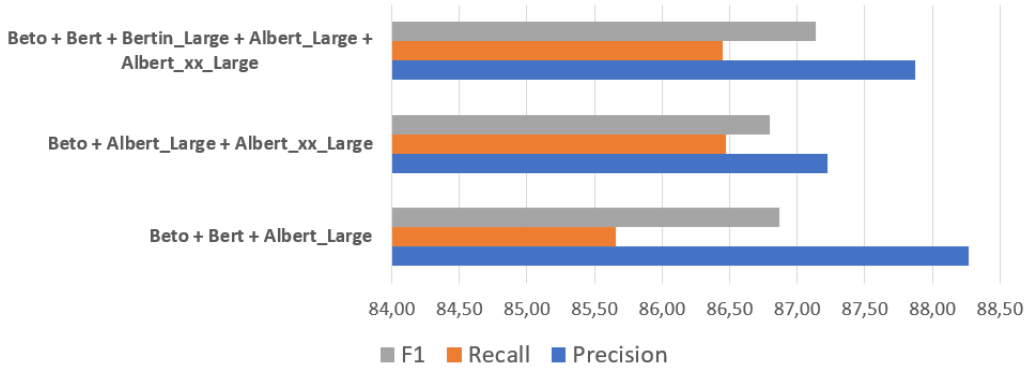


Figure 6. Results of each Transformers combination, independently of the ensemble technique applied.

Table 5: Comparison of the proposed solution with the state of the art.

Id.	**Solutions**	F1-Score
1	GTI/C*	68,51
2	GTI/U*	68,38
3	IIT-T/U*	64,33
4	TGB/C*	55,76
5	*SemEval-2016 – baseline* [11]	51,91
6	[13]	73,07
7	[8]	76,7
8	AspectSA [12]	73,07
9	[21]	73,07
10	[22]	73,0
11	AB1-CNN [7]	65,40
12	BiLSTM-CRF [9]	71,70
13	[23]	71.1
14	[24]	73,1
15	[25]	65,4
16	**Average Voting Ensemble of BETO + BERT + ALBERT_large**	**89,36**

** Solutions submitted to SemEval-2016 Evaluation Workshop in Task 5 (results in Slot2 F-1: ES OTE)* [11].

The superior performance of the ensemble approach over individual transformer models can be attributed to its ability to combine diverse linguistic representations and compensate for the specific biases of each model. Since different pretrained transformers capture complementary semantic patterns and exhibit distinct error behaviours, aggregating their predictions through Average Voting results in more robust and stable outputs. This synergy

enhances generalization to unseen data and reduces the impact of noise or variability in user opinions, ultimately leading to more accurate aspect extraction than any standalone model. The superiority specifically of the BETO + BERT + ALBERT_large ensemble it is the direct result of combining models that are both individually competent and complementary in their errors. The diversity in the training corpus (pure Spanish in BETO vs. multilingual in BERT) and in the architecture (standard BERT vs. ALBERT) creates a cross-verification system where decisions are made collectively, resulting in more accurate, robust, and reliable aspect extraction for user opinions in Spanish.

5 Conclusions and Future Works

In this work, the application of Transformer models and their integration through ensemble learning approaches was studied, in order to identify a solution that improves the effectiveness of aspect extraction in Spanish reviews. As a result, a new aspect extraction method is proposed in which three Transformer models (BETO + BERT + ALBERT-large) are integrated into an ensemble architecture, and the Average Voting technique is used for combining the predictions of the models, and obtaining a final prediction. This solution approach allows us to take advantage of the potentialities of the Transformers-model integrations in the extraction of aspects and constitutes a contribution with respect to other solutions reported in the literature. The experimental evaluation methodology was supported by the use of SemEval2016 dataset in Spanish and included the study and evaluation of 10 individual Transformer models, as well as several integration variants and ensemble techniques. The experimental results show that the application of the Transformer ensemble improves the effectiveness of the aspect extraction task, compared to the use of individual Transformer models and other deep learning techniques. Average Voting is the most promising ensemble technique because it offers a better balance between Precision and Recall, and higher F1. Another important conclusion is that increasing the number of models to be integrated into an ensemble architecture does not necessarily produce a representative improvement of the effectiveness. In future works, the evaluation will be extended to other datasets and other combinations of Transformers will be studied, and aspect extraction methods will be integrated with a process of polarity detection to complete the solution of ABSA. Furthermore, more detailed analyses of the qualities of each model will be carried out to improve the selection of each model to be integrated into the ensemble scheme.

References

[1] B. Liu and L. Zhang, “A Survey of Opinion Mining and Sentiment Analysis,” in Mining Text Data, Boston, MA: Springer US, 2012, pp. 415–463. doi: 10.1007/978-1-4614-3223-4_13.

[2] A. Nazir, Y. Rao, L. Wu, and L. Sun, “Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey,” IEEE Trans. Affect. Comput., vol. 13, no. 2, pp. 845–863, Apr. 2022, doi: 10.1109/TAFFC.2020.2970399.

[3] M. P. Geetha and D. Karthika Renuka, “Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model,” Int. J. Intell. Networks, vol. 2, pp. 64–69, 2021, doi: 10.1016/j.ijin.2021.06.005.

[4] S. U. S. Chebolu, F. Dernoncourt, N. Lipka, and T. Solorio, “A Review of Datasets for Aspect-based Sentiment Analysis,” in Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 611–628. doi: 10.18653/v1/2023.ijcnlp-main.41.

[5] N. Liu, B. Shen, Z. Zhang, Z. Zhang, and K. Mi, “Attention-based Sentiment Reasoner for aspect-based sentiment analysis,” Human-centric Comput. Inf. Sci., vol. 9, no. 1, p. 35, Dec. 2019, doi: 10.1186/s13673-019-0196-3.

[6] D. L. Ramos and L. A. García, “Aprendizaje profundo para la extracción de aspectos en opiniones textuales,” Rev. Cuba. Ciencias Informáticas, vol. 13, no. 2, pp. 105–145, 2019.

[7] B. C. Martínez-Seis, O. Pichardo-Lagunas, S. Miranda, I. J. Perez-Cazares, and J. A. Rodriguez-González, “Deep Learning Approach for Aspect-Based Sentiment Analysis of Restaurants Reviews in Spanish,” Comput. y Sist., vol. 26, no. 2, pp. 899–908, 2022, doi: 10.13053/CyS-26-2-4258.

[8] P. Montañez Castelo, A. Simón-Cuevas, J. A. Olivas, and F. P. Romero, “Enhancing Spanish Aspect-Based Sentiment Analysis Through Deep Learning Approach,” in Progress in Artificial Intelligence and Pattern Recognition. IWAIPR 2023. Lecture Notes in Computer Science, 2024, pp. 215–224. doi: 10.1007/978-3-031-49552-6_19.

[9] T. Chen, R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using

BiLSTM-CRF and CNN," Expert Syst. Appl., vol. 72, pp. 221–230, Apr. 2017, doi: 10.1016/j.eswa.2016.10.065.

[10] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014.

[11] M. Pontiki et al., "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 19–30. doi: 10.18653/v1/S16-1002.

[12] C. Henríquez, F. Briceño, and D. Salcedo, "Unsupervised model for aspect-based sentiment analysis in Spanish," IAENG Int. J. Comput. Sci., vol. 46, no. 3, 2019.

[13] C. Henriquez Miranda and E. Buelvas, "AspectSA: Unsupervised system for aspect based sentiment analysis in Spanish," Prospectiva, vol. 17, no. 1, pp. 87–95, Apr. 2019, doi: 10.15665/rp.v17i1.1961.

[14] A. Mohammadi and A. Shaverizade, "Ensemble Deep Learning for Aspect-based Sentiment Analysis," Int. J. Nonlinear Anal. Appl., vol. 12, no. October 2020, pp. 29–38, 2021.

[15] E.-S. Apostol, A.-G. Pisică, and C.-O. Truică, "ATESA-BERT: A Heterogeneous Ensemble Learning Model for Aspect-Based Sentiment Analysis," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.15920

[16] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Naacl-Hlt 2019, no. Mlm, pp. 4171–4186, 2018, [Online]. Available: https://aclanthology.org/N19-1423.pdf

[17] C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," in Proceedings of International Conference on Learning Representations (ICLR 2020), 2020, pp. 1–18. [Online]. Available: https://github.com/google-research/

[18] J. de la Rosa, E. G. Ponferrada, P. Villegas, P. G. de P. Salas, M. Romero, and M. Grandury, "BERTIN: Efficient Pre-Training of a Spanish Language Model using Perplexity Sampling," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.06814

[19] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:160025533

[20] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," in International Conference on Learning Representations, Sep. 2019. [Online]. Available: http://arxiv.org/abs/1909.11942

[21] C. Henriquez and G. Sanchez-Torres, "Aspect extraction for opinion mining with a semantic model," Eng. Lett., vol. 29, no. 1, pp. 61–67, 2021.

[22] M. S. Akhtar, A. Kumar, A. Ekbal, C. Biemann, and P. Bhattacharyya, "Language-Agnostic Model for Aspect-Based Sentiment Analysis," in Proceedings of the 13th International Conference on Computational Semantics - Long Papers, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 154–164. doi: 10.18653/v1/W19-0413.

[23] M. Vázquez-Hernández, L. Villaseñor-Pineda, and M. Montes-y-Gómez, "A Semantic-Proximity Term-Weighting Scheme for Aspect Category Detection". Procesamiento del Lenguaje Natural, no. 69, 2022, pp. 117-127. doi: 10.26342/2022-69-10.

[24] C. Henriquez, and G. Sánchez-Torres, "Aspect Extraction for Opinion Mining with a Semantic Model". Engineering Letters, vol. 29, no. 1, 2021.

[25] C. Martínez-Seis, O. Pichardo-Lagunas, S. Miranda, Y. Pérez-Cazares, I. J. Caza-res, and J. A. Rodríguez-Gonzalez, "Deep learning approach for aspect-based sentiment analysis of restaurants reviews in Spanish", Computación y Sistemas, nol. 26, vol. 2, 2022, pp. 899-908. doi: 10.13053/CyS-26-2-425.