Overview of FinancES 2023: Financial Targeted Sentiment Analysis in Spanish

Resumen de FinancES 2023: Análisis de Sentimiento Dirigido en Español sobre Finanzas

José Antonio García-Díaz,[1] **Ángela Almela,**[2]
Francisco García-Sánchez,[1] **Gema Alcaraz-Mármol,**[3] **María José Marín,**[2]
Rafael Valencia-García[1]

[1]Facultad de Informática, Universidad de Murcia
[2]Facultad de Letras, Universidad de Murcia
[3]Facultad de Educación, Universidad de Castilla La Mancha
{joseantonio.garcia8, angelalm, frgarcia, mariajose.marin, valencia}@um.es
Gema.Alcaraz@uclm.es

Abstract: This paper presents the FinancES 2023 shared task, organized in the IberLEF 2023 workshop, within the framework of the 39th International Conference of the Spanish Society for Natural Language Processing (SEPLN 2023). The aim of this task is to extend the challenge of sentiment analysis in Spanish to the financial domain, in order to extract the sentiment that a piece of financial information can have for several actors, including the main economic target (i.e., the specific company or asset where the economic fact applies), other companies (i.e., the entities producing the goods and services that others consume) and consumers (i.e., households/individuals). Specifically, two tasks are proposed and evaluated separately. One to identify the main target and to determine the sentiment polarity towards such target, and a second task to assess the sentiment towards both other companies and consumers. The ranking includes results for 10 different teams proposing novel approaches, mostly based on Transformers and generative language models.
Keywords: Sentiment Analysis, Named Entity Recognition, Targeted Sentiment Analysis, Finances.

Resumen: Este artículo resume la tarea FinancES 2023, organizada en el taller IberLEF 2023, dentro del marco de la 39ª Conferencia Internacional de la Sociedad Española de Procesamiento del Lenguaje Natural (SEPLN 2023). El objetivo de esta tarea es mejorar la materia de la minería de opiniones en español dentro del ámbito financiero realizando el análisis de sentimientos desde distintos puntos de vista. En concreto, se proponen y estudian dos tareas que son evaluadas de forma independiente. La primera tarea consiste en (i) identificar el actor principal asociado a una noticia financiera, y (ii) el sentimiento expresado hacia dicho actor. La segunda tarea consiste en determinar el sentimiento de la noticia (i) hacia otras empresas (i.e., otros agentes económicos), y (ii) hacia los consumidores (i.e., la sociedad). El ranking incluye los resultados de 10 equipos diferentes que proponen enfoques novedosos, en su mayoría basados en Transformers y modelos generativos del lenguaje.
Palabras clave: Análisis de Sentimientos, Extracción de Entidades Nombradas, Análisis de Sentimiento Dirigido, Finanzas.

1 Introduction

The need to manage financial data has been in the spotlight for some time (Hasan, Popp, and Oláh, 2020). Years ago, this data sat in warehouses connected to specific applications in banks and financial companies. Then the web came along and made economic, financial and monetary data more accessible. With vast amounts of financial information available online, there are many opportunities to monitor public opinion, receive early warnings and perform positive and negative impact analysis. The impact of emotions on financial markets has been demonstrated in several studies (Goodell et al., 2022; Nemes

and Kiss, 2021).

Several factors hinder the effectiveness of Sentiment Analysis in the financial domain. Firstly, financial language is inherently complex as financial terms refer to an underlying social, economic and legal context (Milne and Chisholm, 2013). Furthermore, the type of language used within this domain is highly context dependent, as a word or phrase can have either positive or negative connotations depending on the context in which it is used (e.g., 'stock market shares rise' vs. 'debt rises'). Finally, the degree of subjectivity in the financial domain is not difficult to determine, as an event can be considered positive or negative depending on the point of view. For example, there may be financial news positive for banks but very negative for other economic sectors and society in general.

The FinancES 2023 shared task is organized at IberLEF 2023 (Jiménez-Zafra, Rangel, and Montes-y Gómez, 2023). The aim of FinancES is to explore targeted Sentiment Analysis in the financial domain. Specifically, the approach adopted here is based on microeconomics. In this context, (Bowles, 2003) explains the role of economic agents, i.e. individuals or organizations that influence the economy. The author states that the main microeconomic agents in the capital market are consumers, companies, governments and central banks.

Consequently, in order to develop a Sentiment Analysis method that takes into account different viewpoints, three different perspectives are considered: (1) economic target of the news item; (2) individual economic actor: companies; and (3) individual economic actor: consumers. The 'target' refers to the sector to which the economic fact applies, while 'companies' are those entities that produce the goods and services that households/individuals consume (i.e., 'consumers'). From these three perspectives, the news item has an impact on the target and the economic agents, which is considered as positive, negative or neutral. To sum up, two distinct tasks are proposed. On the one hand, a task combining the challenges of aspect-term extraction to identify the target entity in the text and aspect-based sentiment classification to determine the emotional polarity towards such target. On the other hand, a task dedicated to assessing the impact of a news headline on the two other economic agents, namely, companies and consumers.

The rest of the paper is organized as follows. Section 2 provides a description of the two tasks involved. Next, the dataset compiled for this task is described in Section 3. The approaches of the participants for solving the two tasks proposed can be found in Section 4. The official leader board is shown and commented in Section 5. Finally, Section 6 summarizes the task and includes some further work.

2 Task description

There are two main challenges in the FinancES 2023 shared task: (i) The identification of the main economic target of the headline, which is made difficult by both the reduced length of the text and the linguistic characteristics of newspaper headlines; and (ii) Multi-dimensional sentiment classification, which, unlike traditional multi-target tasks, where multiple targets are identified within each processed text, here each headline refers to a single target entity, but the stances of other economic agents (companies and consumers) are also taken into account.

Accordingly, two tasks are proposed. The first task, *financial targeted sentiment analysis*, consists of identifying the main economic target from financial news headlines and determining the sentiment polarity (positive, neutral or negative) towards such target. The second task, *financial sentiment analysis at document level for companies and consumers*, consists of determining the sentiment polarity of each news headline towards both companies and consumers.

Task 1 is evaluated with the arithmetic mean of the target f1-score and sentiment classification macro-f1. For the economic target, the IOB2 (short for Inside, Outside, Beginning) scheme is used. Task 2 is evaluated with the arithmetic mean of the macro-F1 for companies and macro-F1 for consumers.

The competition was organized through CodaLab and is accessible at the following link: https://codalab.lisn.upsaclay.fr/competitions/10052. It was organized into 3 phases: Practice, Evaluation and Post-evaluation. In the Practice phase, the participants were provided with a subset of the training data to familiarize with the training data format, and with a notebook with a baseline based on Bag of Words (BoW) and Spacy (Honnibal and Montani, 2017). Later,

| | | | Sentiment | |
text	target	target	companies	consumers
Acuerdos comerciales, sinónimo de oportunidades para República Dominicana	Acuerdos comerciales	POS	POS	POS
EDP Renováveis vende unos activos eólicos en Portugal a China Three Gorges por 242 millones	EDP Renováveis	POS	POS	NEU
El petróleo avista los 82 dólares: la demanda gana a la normalidad en Kazajistán y Libia	Petróleo	POS	NEG	NEG

Table 1: Examples of annotations of the dataset.

the participants were provided with the full training set to develop their approaches. For this, they were allowed to make a maximum of 100 submissions in CodaLab. It should be mentioned that in the Evaluation phase the test partition was provided to participants, who labeled it using the developed systems. This partition was used to evaluate the teams. They were allowed to make a maximum of 10 submissions through CodaLab, from which each team had to select the best one for ranking. The ranking was determined separately for both tasks.

3 Dataset

The dataset for FinancES 2023 is an extension of a the dataset published in (Pan et al., 2023). This dataset consists of Spanish news headlines collected from digital newspapers specialized in economic, financial and political news. Some of these newspapers are Expansión,[1] El Economista,[2] Modaes[3] or El Financiero.[4] It is worth noting that these newspapers are based in different Spanish-speaking countries.

In preparing the dataset, a two-step filtering process was considered. First, specific subsections of the newspapers were identified that contain the news published with economic content (e.g. `elconfidencial.com/mercados`). Furthermore, the collected headlines were pre-processed to discard those that did not belong to the financial domain. Secondly, a content curator manually revised the remaining headlines, removing the irrelevant ones.

Each headline in the corpus was labeled with the target entity and the sentiment polarity on three dimensions: target, companies and consumers. That is, given a headline, it was manually classified as positive, neutral or negative for three specific entities: (i) target (i.e., the specific company or asset to which the economic fact applies), (ii) companies (i.e., the entities that produce the goods and services that others consume), and (iii) consumers (i.e., households/individuals). Each heading was annotated by three members of the organizing committee. In case of disagreement, the annotators discussed the particular case and, if no agreement was reached, the headline was discarded. In this first step, we collected about 14k headlines, filtering out the headlines with a short length or those that did not specify a target entity. Table 1 shows some examples of the dataset, including the headline, the target and the sentiment in a three-class polarity for the target and the other economic agents: companies and society.

The final dataset consists of 7980 headlines, divided in a ratio of almost 80%-20%. It is worth noting that the train split is the same for both tasks. The statistics of the dataset for each sentiment are shown in Table 2. It can be seen that the ratio between positive, neutral and negative documents varies depending on the considered dimension. In the case of the target, there are less headlines that are neutral, while is the other way around for companies and consumers, where there is a greater number of headlines that are considered neutral. On the other hand, the number of negative sentences for all three dimensions is more similar.

---

[1] https://www.expansion.com/

[2] https://www.eleconomista.es/

[3] https://www.modaes.com/

[4] https://www.elfinanciero.com.mx/

José Antonio García-Díaz, Ángela Almela, Francisco García-Sánchez, Gema Alcaraz-Mármol, María José Marín, Rafael Valencia-García

	Target		Companies		Consumers	
label	train	test	train	test	train	test
positive	2818	816	646	523	897	553
neutral	606	205	3843	822	4173	803
negative	2935	600	1870	276	1289	265
total	6359	1621	6359	1621	6359	1621

Table 2: Corpus statistics per sentiment towards the target, companies, and consumers.

4 Participant approaches

There are 34 users registered in the FinancES 2023 shared task. 10 of which submitted results and 6 presented working notes describing their systems. It worth noting that there were two teams that participated in only one of the tasks proposed. Next, we present a brief summary of the participants' proposals:

- **abc111**. This team achieved the 1st position in Task 1 and the 4th position in Task 2. This team did not send the working notes describing their proposed approach.

- **ABCD Team** (Thin et al., 2023). This team achieved the 3rd position in both Task 1 and Task 2. For this, the participants propose a generative framework based on the mT5 model (Xue et al., 2021) that addresses both tasks. To support the model in distinguishing sentiment values for the two tasks, the authors introduce two binary auxiliary tasks that aim to (i) identify whether the sentiment values of Task 1 and Task 2 align, and (ii) determine whether the sentiment values between the companies and consumers are equivalent.

- **AnkitSinghRaikuni**. This team achieved the 5th position in the Task 1 and the 6th position in Task 2. This team did not send the working notes describing their proposed approach.

- **ITST** (Salas-Zárate and Paredes-Valverde, 2023). This team achieved the 8th position in Task 1 and the 9th position in Task 2, both below the proposed baseline. For this, the participants present a system based on linguistic features which are extracted with LIWC (Tausczik and Pennebaker, 2010). Since some of the identified features might not be relevant to the financial domain,

the features are discriminated using a selection technique based on Rough Set Theory and Information Gain. Overall, the resulting performance is limited as compared to state-of-the-art approaches.

- **NLP_URJC** (Rodríguez-García et al., 2023). This team achieved the 7th position in both tasks; however, they only outperformed the proposed baseline in the case of Task 1. Their pipeline consists of a preprocessing stage, an entity recognition phase and the polarity extraction. All these models are grounded on the usage of BETO (Cañete et al., 2020) and spaCy (Honnibal and Montani, 2017) for the part-of-speech tagging.

- **LLI-UAM** (Laboratorio de Lingüística Informática) (Porta-Zamorano, Torterolo, and Moreno-Sandoval, 2023). This team achieved the 2nd position in Task 1 and the 1st position in Task 2. This team evaluates Transformers based on different sizes of the mT5 model (after fine-tuning) for sentiment classification. They also conduct a data analysis process for identifying and correcting mistakes in targets and the polarity. An interesting approach proposed by this team is the usage of ChatGPT for data augmentation to increase the datasets' distribution.

- **SINAI** (Jiménez-Zafra et al., 2023). The SINAI team achieved the 4th position in Task 1 and the 2nd position in Task 2. They explored Transformer-based models pretrained with general data and financial domain data as well as off-the-shelf NER systems. The models are monolingual in Spanish and English, translating the dataset into English as a preliminary stage to evalu-

Ranking	User	Result	Target	Sentiment
01	abc111	79.224	87.714	70.735
02	LLI-UAM	79.217	85.218	73.216
03	ABCD Team	78.218	85.451	70.984
04	SINAI	77.800	83.817	71.783
05	Ankit Singh Raikuni	55.421	57.536	53.306
06	UTB-NLP	52.923	41.008	64.838
07	NLP_URJC	51.441	60.677	42.206
–	BASELINE	49.811	42.839	56.782
08	Team ITST	27.693	10.633	44.753
09	UNAM Text Mining	13.468	08.664	18.272

Table 3: FinancES official leaderboard for Task 1, including the official ranking and result. We also provide the results for identifying the target and the sentiment.

Ranking	User	Result	Companies	Consumers
01	LLI-UAM	64.235	59.259	69.211
02	SINAI	63.490	58.348	68.632
03	ABCD Team	61.037	58.864	63.211
04	abc111	57.502	53.028	61.975
05	fanchuyi	47.269	41.423	53.114
06	Ankit Singh Raikuni	45.763	41.976	49.551
–	BASELINE	43.378	38.427	48.330
07	NLP_URJC	42.513	43.656	41.369
08	UNAM Text Mining	37.040	34.569	39.511
09	Team ITST	24.820	26.927	22.713

Table 4: FinancES official leaderboard for Task 2, including the official ranking and result. We also provide the results for identifying the sentiments for the companies and consumers.

ate the English-based Transformers. Besides, they evaluated generative LLMs such as ChatGPT.

- **UNAM Text Mining**. This team achieved the 9th position in Task 1 and the 8th position in Task 2, both below the proposed baseline. This team did not send the working notes describing their proposed approach.

- **UTB-NLP** (Cuadrado et al., 2023). This team only participated in Task 1, achieving the 6th position. This task is addressed by combining Transformer-based models and phonestheme embeddings, feeding the extracted features into a traditional machine learning classifier based on Support Vector Machines.

- **fanchuyi**. This team only participated in Task 2, achieving the 5th position. This team did not send the working notes describing their proposed approach.

5   Results and discussion

The official leaderboards of the FinancES 2023 shared task are shown in Table 3 for Task 1 and in Table 4 for Task 2.

For Task 1 (see Table 3) the best result is achieved by *abc111*, with an average score of 79.224%. This result is very similar to that of the *LLI-UAM* team (79.217%). The main difference is that *LLI-UAM* scored better in identifying the sentiment and *abc111* scored better in identifying the target. It can be seen that the top four teams achieve similar results in terms of identifying the target and the sentiment, but there is a significant drop in performance in the results from position 5 onwards. The low results obtained by the *UNAM Text Mining* team in identifying the target suggest a problem with their pipeline, as they obtain more limited results than those obtained with a random classifier. In this case, all but two participants beat the proposed baseline of spaCy for target identification and BoW for sentiment classification.

For Task 2 (see Table 4), the best result is achieved by *LLI-UAM* with a score of 64.235%, less than one point above *SINAI* (the second ranked team with a score of 63.490%). It can be seen that the results are generally better for identifying sentiment towards consumers than towards other companies. In this case, the drop in performance is also observed between the fourth best participant and the rest. The baseline, consisting of a BoW, outperformed three participants for both company and consumer sentiment identification.

6   Conclusions

This paper presents the first edition of the FinancES shared-task in IberLEF 2023, which consists of two subtasks: one to identify the main target of a financial news headline and its sentiment, and another to calculate the sentiment polarity towards consumers and other companies.

As this is the first time we have organized this event, we are very pleased with the response, with the registration of 34 users and the participation of 10 teams, who sent promising approaches for solving both tasks, most based on Transformers and generative models.

The shared task is still accessible in the post-evaluation phase `https://codalab.lisn.upsaclay.fr/competitions/10052`. The Codalab page contains two notebooks for the preparation of the baseline and the submission file to be sent to the competition, as well as the full dataset including the golden labels. We hope that these resources will help the Spanish NLP community.

As future work, we would like to organize a second edition of this task. In this sense, we will extend the dataset incorporating more data sources, including more newspapers and other sources such as micro-posts from Twitter. Incorporating more subjective content, we are considering to annotate the dataset from a Emotion Analysis perspective in order to incorporate a more fine-grained subjective analysis of financial news.

Acknowledgements

References

Bowles, S. 2003. *Microeconomics: behavior, institutions, and evolution.* Princeton University Press.

Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.

Cuadrado, J., E. Martinez, J. C. Martinez-Santos, and E. Puertas. 2023. Team UTB-NLP at FinancES 2023: Financial Targeted Sentiment Analysis Using a Phonestheme Semantic Approach. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Goodell, J. W., S. Kumar, P. Rao, and S. Verma. 2022. Emotions and stock market anomalies: a systematic review. *Journal of Behavioral and Experimental Finance*, page 100722.

Hasan, M. M., J. Popp, and J. Oláh. 2020. Current landscape and influence of big data on finance. *Journal of Big Data*, 7(1):1–17.

Honnibal, M. and I. Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Jiménez-Zafra, S. M., D. García-Baena, M. Ángel García-Cumbreras, and M. García-Vega. 2023. SINAI at FinancESIberLEF2023: Evaluating Popular Tools and Transformers Models for Financial Target Detection and Sentiment Analysis. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Jiménez-Zafra, S. M., F. Rangel, and M. Montes-y Gómez. 2023. Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Milne, A. and M. Chisholm. 2013. The prospects for common financial language in wholesale financial services. *SSRN Electronic Journal.*

Nemes, L. and A. Kiss. 2021. Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 5(3):375–394.

Pan, R., J. A. García-Díaz, F. Garcia-Sanchez, and R. Valencia-García. 2023. Evaluation of transformer models for financial targeted sentiment analysis in spanish. *PeerJ Computer Science*, 9:e1377.

Porta-Zamorano, J., Y. Torterolo, and A. Moreno-Sandoval. 2023. LLI-UAM Team at FinancES 2023: Noise, Data Augmentation and Hallucinations. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Rodríguez-García, M. A., A. Riaño-Martínez, D. Roldán-Álvarez, and S. Montalvo-Herranz. 2023. URJC-Team at FinancES 2023: Financial Targeted Sentiment Analysis in Spanish Combining Transformers. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Salas-Zárate, M. d. P. and M. A. Paredes-Valverde. 2023. Team ITST at FinancES 2023: A Psycholinguistic-based Sentiment Analysis Approach. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Tausczik, Y. R. and J. W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Thin, D. V., D. N. Ba, D. N. Hao, and N. L.-T. Nguyen. 2023. ABCD Team at FinancES 2023: An Unified Generative Framework for the Financial Targeted Sentiment Analysis in Spanish. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org.*

Xue, L., N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tür, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 483–498. Association for Computational Linguistics.