

# INVESTIGATING THE RELATIONSHIP BETWEEN ROMANIAN FINANCIAL NEWS AND CLOSING PRICES FROM THE BUCHAREST STOCK EXCHANGE

Ioan-Bogdan Iordache<sup>1,2</sup>, Ana Sabina Uban<sup>1,2</sup>, Catalin Stoean<sup>1,3</sup>, Liviu P. Dinu<sup>1,2</sup>

<sup>1</sup>Human Language Technology Research Center, Bucharest, Romania,

<sup>2</sup>Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest, Romania,

<sup>3</sup>Faculty of Sciences, University of Craiova, Craiova, Romania



## Abstract

A new data set is gathered from a Romanian financial news website for the duration of four years. It is further refined to extract only information related to one company by selecting only paragraphs and even sentences that referred to it. The relation between the extracted sentiment scores of the texts and the stock prices from the corresponding dates is investigated using various approaches like the lexicon-based Vader tool, Financial BERT, as well as Transformer-based models. Automated translation is used, since some models could be only applied for texts in English. It is encouraging that all models, be that they are applied to Romanian or English texts, indicate a correlation between the sentiment scores and the increase or decrease of the stock closing prices.

## Introduction

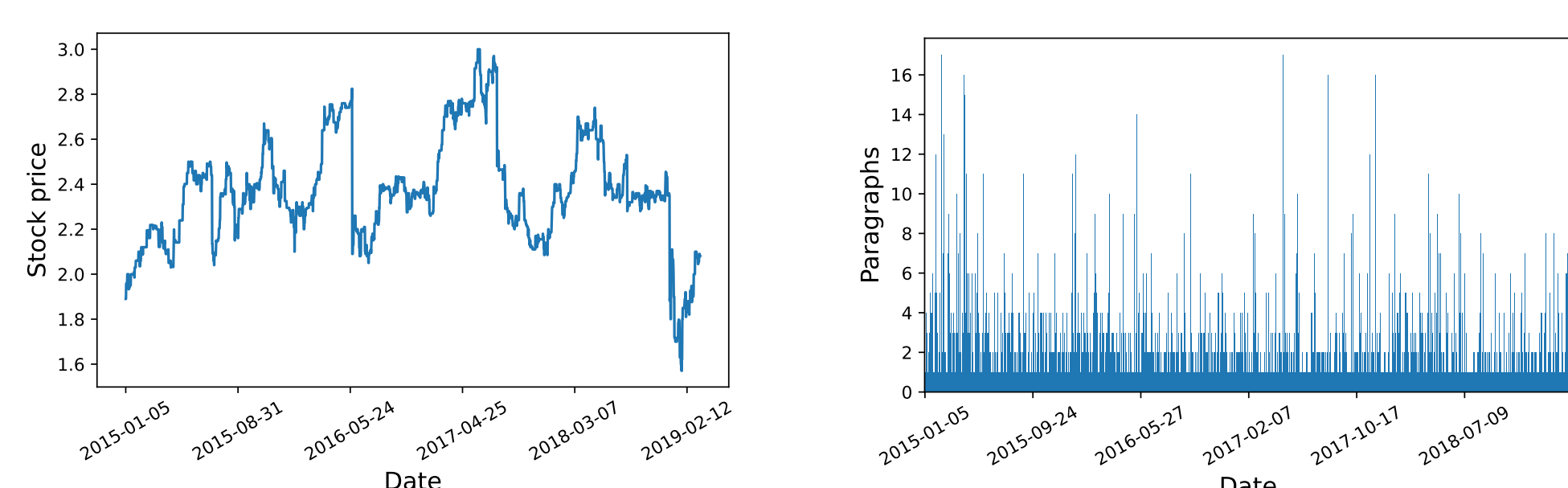
In the current work, we focus on one company and we analyse the information that appears in the press for verifying the correlation between sentiments extracted from the texts and the closing prices of the company.

The dataset and the methods we provide are tailored for the Romanian language. Although a work in progress, the identified statistics encourage us to further investigate means of using the extracted sentiment scores for empowering stock price prediction.

## Dataset

The dataset is gathered from online articles published between **January 5, 2015** and **April 25, 2019** by the newspaper Bursa which can be accessed at bursa.ro.

We decided to focus on one company, **Transilvania Bank** (ranked first in Romania based on bank assets). Articles mentioning this company were selected ( $\approx 4.36\%$  of all articles). Since the statements connected to the different companies often had distinct, even contradictory sentiments, we limited our dataset to the **paragraphs** that mention Transilvania Bank. We also alternatively reduced the paragraphs to **sentences**.



(a) Closing prices for Transilvania Bank for the entire period (b) Number of paragraphs extracted over time

Fig. 1: Closing prices and article occurrences of Transilvania Bank.

## Experiments

Since most available resources are developed for English, we include approaches where we first translate the Romanian texts in our dataset into English.

**Lexicon-based sentiment scores.** From NLTK library, the Vader (Valence Aware Dictionary and sEntiment Reasoner) [HG14] tool is used for extracting scores for the texts in English. Vader outputs 4 different scores, i.e. positive, negative, neutral and compound.

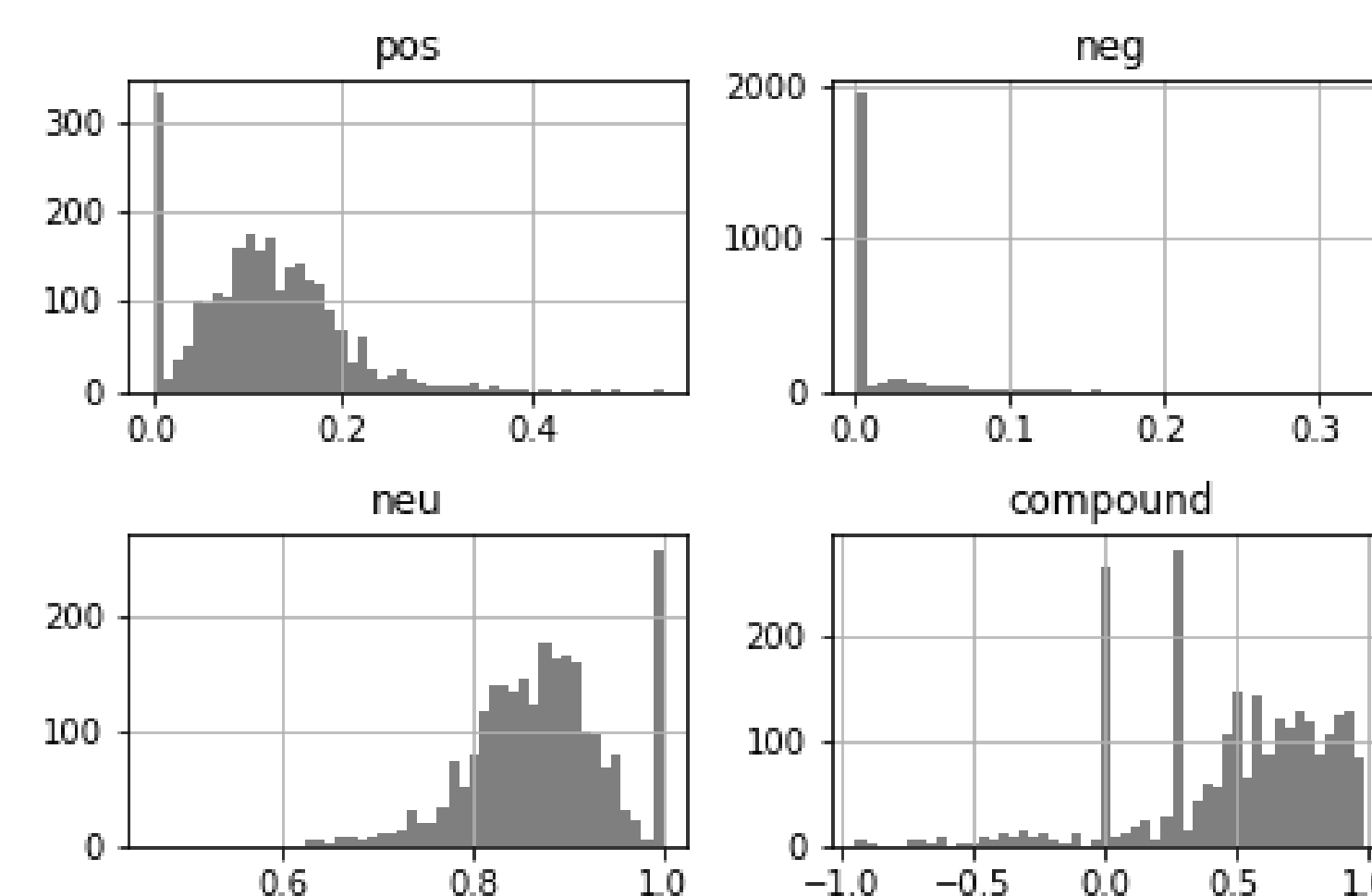


Fig. 2: Histogram for the values of the positive, negative, neutral and compound scores obtained using Vader.

**Financial BERT.** In order to provide sentiment scores for the paragraphs and sentences that were translated into English, we employ a BERT model that was fine-tuned for the task of sentiment analysis on financial texts (FinBERT) [Ara19].

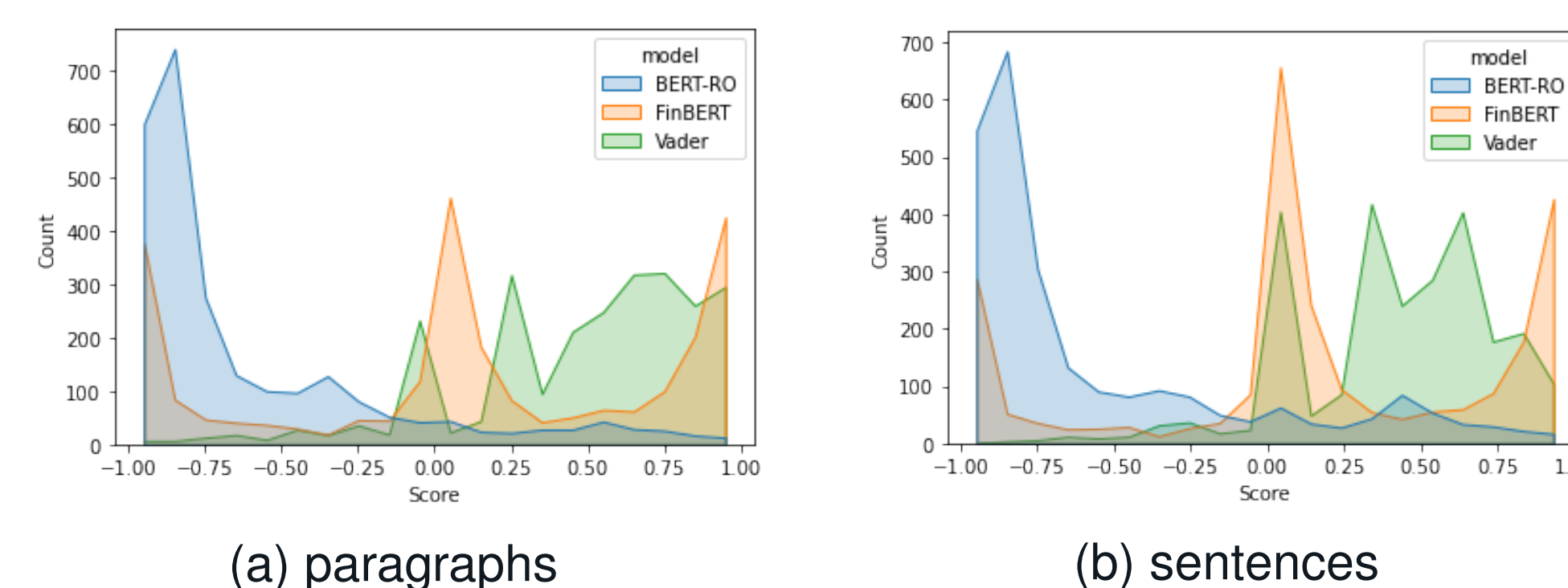


Fig. 3: Sentiment score distribution for both Romanian and English paragraphs and sentences using BERT-based and lexicon-based metrics.

**Transfer learning from Romanian sentiment-annotated corpus.** For scoring Romanian texts, we trained a BERT model [Dev+18] on a large dataset of reviews from Romanian e-commerce websites (LaRoSeDa) [TGI21]. The classification confidence of the model is used for determining the sentiment score.

	positive	negative	neutral
Transfer Paragraph-RO	10.88%	89.12%	—
Transfer Sentence-RO	16.24%	83.76%	—
FinBERT Paragraph-EN	36.64%	26.32%	37.04%
FinBERT Sentence-EN	33.51%	19.61%	46.88%

Tab. 1: Distribution of the labels predicted for the financial texts (paragraphs and sentences) from our dataset by the BERT-based models, trained for sentiment analysis on Romanian texts (with positive and negative labels) and English texts (with positive, negative, and neutral labels).

## Sentiment and Stock Price

In order to assess the potential influence of the sentiment expressed in the media on the stock price, we compare the evolution of the sentiment scores over time with the evolution of the stock price, using **Pearson correlation**.

Sentiment scores are correlated with stock price: the positive sentiment is correlated with the price, and the negative sentiment is anti-correlated with the stock price (for metrics where a separate score is available for the negative sentiment), for all metrics considered.

Metric	Correlation		
	positive	negative	neutral
Vader Sentence-EN	0.050*	-0.031	-0.039
Vader Paragraph-EN	0.057*	-0.040	-0.012*
FinBERT Sentence-EN	-0.003	-0.048*	0.045*
FinBERT Paragraph-EN	0.035	-0.044*	0.009*
Transfer Sentence-RO	0.044*	—	—
Transfer Paragraph-RO	0.054*	—	—

Tab. 2: Correlations between sentiment scores and stock price for the different sentiment metrics at sentence and paragraph level, separately for the positive, negative and neutral sentiment scores (where available).

\*statistically significant ( $p < 0.05$ )

While the correlations are not very strong, the consistent pattern across the different metrics, which persists on the English translations of the texts, suggests that we can indeed find an effect which connects sentiment with stock price.

## Conclusions

- We have built a dataset extracted from Romanian financial news and we proposed methods for assessing the sentiment of the texts;
- In the future, we intend to extend our analysis at the sentiment level to exploring **emotions** expressed in the text and their relationship with the price;
- Additionally, we suggest that rather than positive or negative sentiment, **optimism or pessimism** could be a more suitable predictor of stock price.
- We also intend to complement the correlation analysis between sentiment and current stock price with predictive methods.

## References

- [Ara19] Dogu Araci. "Finbert: Financial sentiment analysis with pre-trained language models". In: *arXiv preprint arXiv:1908.10063* (2019).
- [Dev+18] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).
- [HG14] Clayton J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". In: *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*. Ed. by Eytan Adar et al. The AAAI Press, 2014. URL: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109>.
- [TGI21] Anca Maria Tache, Mihaela Gaman, and Radu Tudor Ionescu. "Clustering Word Embeddings with Self-Organizing Maps. Application on LaRoSeDa—A Large Romanian Sentiment Data Set". In: *arXiv preprint arXiv:2101.04197* (2021).