

# 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, University of Bucharest, 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,

iordache.bogdan1998@gmail.com, auban@fmi.unibuc.ro, catalin.stoean@inf.ucv.ro, ldinu@fmi.unibuc.ro

## 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.

**Keywords:** sentiment analysis, low resource languages, transformers, stock exchange

## 1. Introduction

The prediction of the stock prices and, more recently, the one of the cryptocurrencies, represents an important challenge for economists and machine learning scientists. The efforts are generally focused on designing models that identify specific patterns that determine the changes in the stock price market and accordingly predict the value for the next hour, day, month or even year. The prediction is performed for each company separately and, at a time  $t$ , it is often based on the price stock from the previous  $N$  days, as well as for the number of transactions that were achieved, the amount of shares, the minimum and maximum prices for them in that period, the opening price and so on (Rouf et al., 2021; Gandhmal and Kumar, 2019; Shah et al., 2019).

Relying the prediction only on previous information about the transactions of stocks is clearly not enough. A model that seems to perform well for the stocks of a certain company for predicting its closing price for a period of time does not lead to similarly good outputs on a different time frame. Additionally, a model that appears to be suitable for one company lead to modest results for a different one. In conclusion, designing a model that is based only on numerical information about transactions resembles to a Sisyphian effort. In the current work, we focus on only one company and we analyse the information that appears in the press regarding that company for verifying the sentiments extracted from the texts and the closing prices of the company. Plus, we use texts in Romanian, fact that further complicates the task, since most of the tools for sentiment analysis are dedicated to the English language and the options for Romanian are rather scarce. The current study represents a work in progress, since herein we analyse the sentiment from texts and only verify whether it is correlated to the closing price for

the shares of that stock, without further applying any model for time series prediction with the aim for indicating future values for the closing price. Still, the found results demonstrate that the sentiment scores are correlated with the stock closing price, fact that encourages us to further investigate means of using the extracted sentiment scores for empowering prediction models with such information and later help users decide major buy or sell actions.

## 2. Previous Work

The use of information based of statistics from previous data for determining next closing price for stocks represents a field that has been thoroughly researched, sometimes with encouraging results, but most of the time the overall conclusion was that such data is not enough for reaching informed decisions regarding whether to buy or sell shares (Rouf et al., 2021; Gandhmal and Kumar, 2019; Shah et al., 2019). One previous such study where developed models were applied on a dataset from the Romanian stock exchange is presented in (Stoean et al., 2019). However, this is not the goal of the current article, therefore, the focus will be set of the extraction of sentiment scores from articles. The aim herein is to evaluate if such information related to sentiment scores from articles, extracted from text translated from Romanian to English, can be correlated with the changes in the stock price of a company.

There are many studies that have the goal to predict stock prices based on the sentiments extracted from related news. However, most of the works are based on texts written in the English language. In (Wan et al., 2021), 87 companies are watched in the period 2007-2013, by following news articles from Reuters. The propagation of sentiments extracted via NLP techniques is evaluated with respect to stock price and

volatility. The study (Nemes and Kiss, 2021) applies several tools like BERT, VADER, TextBlob, and a Recurrent Neural Network on headlines of economic news and afterwards compares the sentiment outputs with the stock price moves. In another work (Kollintza-Kyriakoulia et al., 2018), positive and negative opinions from Twitter posts related to five companies on the US market have been correlated with their historical price movement for 4 months of 2015 by means of symbolic aggregate approximation and dynamic time warping.

The use of sentiment analysis extracted from news for stock price prediction when dealing with non-English languages is rather limited. In fact, it is hard to find large annotated corpora for non-English languages in general, hence the extraction of sentiment in general (mainly associated with social media) from non-English texts is relatively scarce, but there are some entries. For instance, Agüero-Torales et al. (2021) presents an excellent survey of deep learning approaches for multilingual sentiment analysis from social media. There are also tools that have a direct interaction with English for reaching the sentiment scores associated to the texts. One such example is (Kaity and Balakrishnan, 2019), where words are extracted from an Arabic corpus and polarity scores are searched for them in an English lexicon, forming this way an Arabic lexicon. Another work (Barriere and Balahur, 2020) pre-trains a multilingual transformer model on English tweets and uses data-augmentation via automatic translation for making the approach suitable for non-English languages. The experiments on French, Spanish, German and Italian proved to be efficient, according to the authors. In (Li et al., 2020), four sentiment dictionaries are tried together with a LSTM model that also incorporates technical indicators for the Hong Kong stock market and the one that led to the most accurate results is the domain specific Loughran-McDonald financial dictionary.

In a previous attempt to use data from the same source as in the current work, the study (Stoean and Lichtblau, 2021) puts forward a method that uses chaos game representation to encode text into grayscale images and then uses the obtained representations with a neural network to link them with the sentiments of the initial articles. The approach was previously used with success for authorship attribution (Lichtblau and Stoean, 2018; Stoean and Lichtblau, 2020), where the authors of the texts were connected to the images that encoded their texts. For having ground truth scores for the sentiments in the articles, automated translation was used and then Vader was used for getting positive, negative, neutral and compound values. The approach involving chaos game representation was applied directly on the texts in Romanian and the results compared favourably with the results provided by Vader.

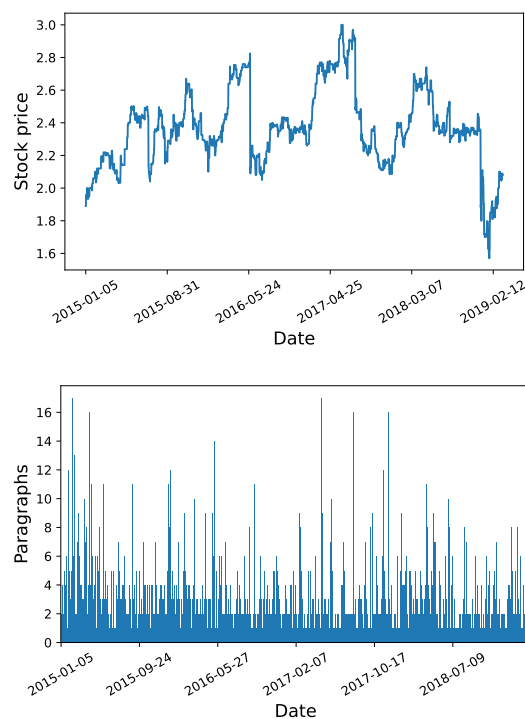


Figure 1: Closing prices for Transilvania Bank for the entire period, on the top plot, and the number of paragraphs about the companies that are extracted over time, on the bottom plot.

### 3. Dataset Acquisition

The dataset is gathered from the online version of the newspaper Bursa which can be accessed at bursa.ro. All articles from January 5, 2015 until April 25, 2019 were gathered, which led to 39731 distinct items. Out of these, we decided to focus on only one company, Transilvania Bank, which has been ranked first in Romania since 2018 till present, based on the bank assets. The aim on the long term is to find proper means to correlate the news findings with the changes in the stock prices and accordingly to be able to anticipate such fluctuations. In this sense, for tracking the connection between prices and related news, we decided to extract information related to a specific company.

The selection of the relevant articles was straightforward, since we filtered only the articles that contained at least one occurrence of the text "Banca Transilvania" (in English, "Transilvania Bank"). This led to 1732 articles, which represents 4.36% of the initial dataset.

In a subsequent reading of several random articles from the obtained dataset, we noticed that many of these texts contained data about several companies, including Transilvania Bank. Moreover, the statements connected to the different companies often had distinct, even contradictory sentiments. Accordingly, we reduced the texts to the paragraphs in which the specific string appeared. Some of the articles have several paragraphs containing the given term, hence, although the

size of each separate text is substantially reduced, the number of instances is now 2519. Their sizes vary from a minimum of 52 characters to a maximum of 2058, while having an average of 394.

In an attempt to being even more restrictive and have the sentiment more precisely connected to the stock of the company, we alternatively reduced the paragraphs to sentences. For these, the same minimum size as for paragraphs is maintained, but the maximum is of 1455 characters, while the average is decreased to 219 characters.

Figure 1 illustrates the changes in the stock price for the entire period in which the news articles are collected, as well as the number of paragraphs related to the companies that are used in the current study.

## 4. Experiments

We use several different approaches for extracting sentiment scores from the texts in our dataset. Since the data is not annotated with sentiment, we use external resources in order to infer the sentiment in texts, including sentiment lexicons, pre-trained models, as well as an additional annotated dataset.

### 4.1. Methods for Sentiment Analysis

Since most available resources are developed for English, we include approaches where we first translate the Romanian texts in our dataset into English using the *deep\_translator* library, version 1.4.4 in Python, then compute sentiment scores on the English texts in order to infer the sentiment in the original Romanian texts. Finally, we also include a metric which can be applied directly on the Romanian texts, relying on transfer learning from an external corpus of Romanian texts annotated for opinion mining.

#### 4.1.1. Lexicon-based sentiment scores

From NLTK library (Bird et al., 2009), the Vader (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014) tool is next used for extracting scores for the texts in English. It is a rule-based sentiment analyzer that uses lexical features labeled as positive or negative to calculate sentiment score for the whole text. The model was particularly designed for social media. Vader outputs 4 different scores, i.e. positive, negative, neutral and compound. While for the first three the values are within the interval  $[0, 1]$ , the scores of the latter are in  $[-1, 1]$ . Figure 2 illustrates histograms for the positive, negative, neutral and compound scores. Two samples of paragraphs in Romanian and English, beside their compound values, are illustrated in Table 1. The samples are chosen to illustrate an example of a positive sentiment and one labeled as negative by the compound score.

#### 4.1.2. Financial BERT

Neural network architectures based on Transformers (Vaswani et al., 2017) have sparked great interest in

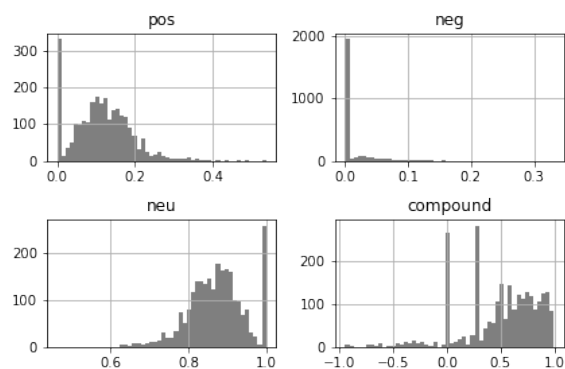


Figure 2: Histogram for the values of the positive, negative, neutral and compound scores obtained using Vader.

the field of Natural Language Processing, by providing state-of-the-art results on a large variety of tasks. Among these architectures, the most notable ones are the models trained on large corpora in a self-supervised manner, using techniques such as Masked-Language Modeling and Next Sentence Prediction. Using the pre-trained weights and the knowledge about language patterns, inferred from the large datasets, these models can be then fine-tuned on downstream tasks, using reasonably-sized corpora and they can display great performance. One such model is BERT (Devlin et al., 2018).

In order to provide sentiment scores for the paragraphs and sentences that were translated into English from our dataset, we employ a BERT-based model that was fine-tuned for the task of sentiment analysis on financial texts (FinBERT) (Araci, 2019). The model's implementation and its weights were acquired through the Huggingface library (Wolf et al., 2019). As the last layer of the architecture, a classification head is provided, that can be used to estimate the confidence that a given text is attributed a positive, neutral or negative sentiment. The confidence values are positive numbers that add up to 1, so they can be used as percentages.

In terms of pre-processing, we tokenize our English texts using a BERT-specific tokenizer and we limited the length of the token sequences to 300. If for a text the number of tokens exceeds this limit, we only keep the first 300 and truncate the rest. Special tokens for the beginning and the end of the sequence are also added. This limit is large enough for almost all of our paragraphs and sentences (Figures 3 and 4). Using the predictions made by the FinBERT model, we provide for each text a sentiment score equal to the difference between the confidence for the positive class and the confidence for the negative class, such that very positive examples will be scored with a value close to 1, while very negative examples will approach the value  $-1$ .

Paragraph in Romanian	Paragraph in English	Compound
In timp ce bancile straine se retrag din Romania sau sunt cumparate, o banca romaneasca, Banca Transilvania, se dezvolta atat in tara cat si in strainatate. Dupa preluarea Volksbank, BT are planuri sa se extinda in Italia, unde exista o comunitate puternica de romani, datorita succesului inregistrat de sucursala deschisa exact in urma cu un an, la Roma. Astfel, de la inaugurarea operatiunilor sucursalei BT din Roma si pana in prezent, aceasta a reusit sa atraga un numar de 1.200 de clienti, care au 1.000 de carduri si 100 de depozite, conform datelor furnizate de banca. Numarul transferurilor depaseste 300/luna iar tendinta este de crestere accentuata, dat fiind faptul ca, din noiembrie 2014, BT Italia a devenit membru SEPA, asa incat BT realizeaza incasari si plati in euro in conditii foarte avantajoase pentru clienti, precizeaza banca. De asemenea, Banca Transilvania - Sucursala Italia este in curs de implementare a produselor de creditare pentru romanii care locuiesc si lucreaza in Italia. Estimarea este ca in primul trimestru al acestui an vor fi acordate primele credite.	While foreign banks withdraw from Romania or are bought, a Romanian bank, Banca Transilvania, is developing both in the country and abroad. After taking over Volksbank, BT plans to expand in Italy, where there is a strong Romanian community, due to the success of the branch opened exactly one year ago in Rome. Thus, since the inauguration of the BT branch operations in Rome and until now, it has managed to attract a number of 1,200 customers, who have 1,000 cards and 100 deposits, according to the data provided by the bank. The number of transfers exceeds 300 / month and the trend is accentuated, given the fact that, since November 2014, BT Italia has become a SEPA member, so BT makes receipts and payments in euros in very advantageous conditions for customers, says the bank. Also, Banca Transilvania - Italy Branch is in the process of implementing credit products for Romanians living and working in Italy. The estimate is that in the first quarter of this year the first loans will be granted.	0.957
Lovitură pentru cei cu credite în franci, după ce cursul BNR a crescut ieri cu peste 15% până la 4,33 lei pentru un franc. Șocul survine după ce Banca Elveției a renunțat la pragul de cotare de 1,20 euro pentru un franc. Circa 150.000 de români mai au de returnat băncilor echivalentul a aproape 10 miliarde lei. După ce banca centrală a Elveției a renunțat la pragul minim de 1,2 franci pe euro, francul elvețian a sărit practic în aer. Ieri la prânz, timp de câteva zeci de minute, mai multe bănci comerciale au afișat un curs peste 5 lei pentru un franc elvețian. Leumi Bank, Banca Transilvania și Bancpost au afișat cursuri peste 5 lei pentru un franc, ultima dintre ele stabilind la un moment dat un curs de schimb de 6,5 lei/franc elvețian.	A blow for those with loans in francs, after the BNR exchange rate increased yesterday by over 15% to 4.33 lei for one franc. The shock comes after the Bank of Switzerland dropped the quotation threshold of 1.20 euros for one franc. About 150,000 Romanians still have to return to the banks the equivalent of almost 10 billion lei. After Switzerland's central bank dropped its threshold of 1.2 francs per euro, the Swiss franc virtually blew up. Yesterday at noon, for several tens of minutes, several commercial banks posted an exchange rate of over 5 lei for a Swiss franc. Leumi Bank, Banca Transilvania and Bancpost posted exchange rates over 5 lei for one franc, the last of which at one point set an exchange rate of 6.5 lei / Swiss franc.	-0.128

Table 1: Two samples of paragraph instances in Romanian (first column), English using automated translation (second column), and the compound score as attributed by Vader for the English version in the third column.

#### 4.1.3. Transfer learning from Romanian sentiment-annotated corpus

For scoring the Romanian paragraphs and sentences from our dataset using Transformer-based models, we took a slightly different approach. Due to the scarcity of annotated Romanian datasets for sentiment analysis in general, not only for financial texts, we ended up with employing one of the more recent and larger resources, the LaRoSeDa dataset (Tache et al., 2021). This dataset contains user reviews gathered from various Romanian e-commerce websites, and were automatically annotated using the scores provided by the

users on these platforms. A total of 15,000 reviews were annotated with one of two labels, corresponding to the sentiment orientation (positive or negative). The text of a review is split into the title of the review and its body.

Using the weights of a 12-layer BERT model, pre-trained on large Romanian corpora (Dumitrescu et al., 2020), we build an architecture for classifying the sentiment of these reviews. This is done by appending a classification head with two output nodes and *softmax* activation after the BERT model. The input of the classification head is the hidden state computed for the ini-

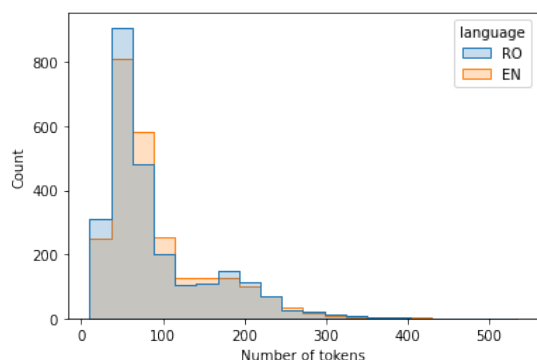


Figure 3: Distribution of the length of the **paragraphs** from our financial texts dataset, in both Romanian and English, computed as the number of tokens outputted by the BERT tokenizer.

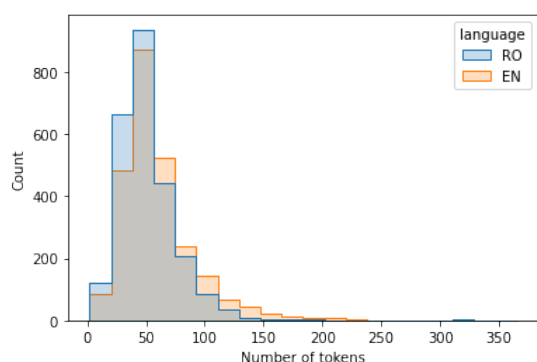


Figure 4: Distribution of the length of the **sentences** from our financial texts dataset, in both Romanian and English, computed as the number of tokens outputted by the BERT tokenizer.

tial token of the sequence (the special token [CLS]), on which 0.1 dropout rate is applied during training. For pre-processing, the text of the title and the text of the body were concatenated for each of the reviews in the dataset. The resulting text is tokenized using the tokenizer provided for the Romanian BERT model and, as before, the same limit of 300 tokens was applied. We used the same split proposed by the authors of the LaRoSeDa dataset, 80% of the dataset is used for training the architecture, while the rest of 20% is used for validation. The model is trained end-to-end using the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.00002 (used for fine-tuning the pre-trained weights) and no weight decay. The loss function of choice is the cross entropy loss. The training examples are provided in mini-batches of size 16, and the training is done for a total of 5 epochs. At the end of each epoch, we use the validation examples in order to compute the performance of our model on unseen data. Based on that, we store the weights of the model that obtains the best accuracy on the validation set. We observed that the model converges after about 4 – 5 epochs, with a validation accuracy of 97.43%.

The model trained using the described strategy is then employed for assigning sentiment scores for our financial texts. Using the same pre-processing technique and passing the texts through the architecture, we can compute the model’s confidence for each label (positive or negative), as positive numbers that add up to 1. The final score for a sentence or paragraph is the difference between the positive label confidence and the negative one, such that, just as before, for very positive texts the score approaches 1, while for very negative ones it approaches  $-1$ .

Using the scores computed by both the English and the Romanian models, we looked at their distribution for both paragraphs and sentences (Figures 5 and 6). The scores attributed by the English FinBERT model are somewhat balanced towards the positive, neutral and negative classes, while the Romanian model’s predictions are slightly biased towards the negative sentiments. The clear disadvantage of the Romanian model is the fact that it was trained on data coming from a different domain (product reviews instead of financial news texts), but as we will later show, the scores it predicts can still be positively correlated with the registered market changes. Table 2 shows the distribution of our financial texts with respect to the labels defined by our BERT-based models. By looking at the texts classified as negative by the English model, we can see that 98.53% of them were classified as negative by the Romanian model as well, while only 19.07% of the positively classified English texts were also labelled by the Romanian model with a positive sentiment.

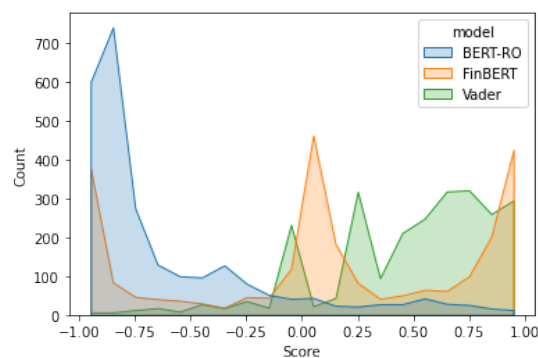


Figure 5: Sentiment score distribution for both Romanian and English **paragraphs** using BERT-based and lexicon-based metrics.

## 4.2. 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. We compute the correlation between each sentiment metric for texts at a certain point of time with the stock price at that point in time, at a daily level (for the same day). Using correlation to measure the connection between sentiment score and stock price has also

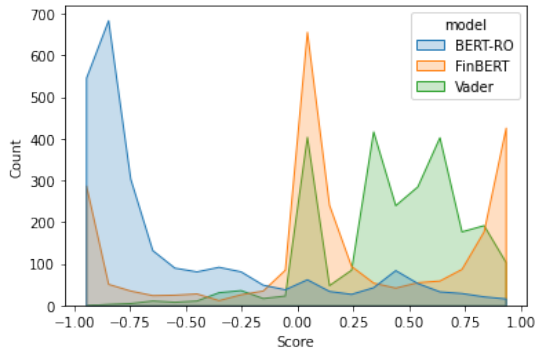


Figure 6: Sentiment score distribution for both Romanian and English **sentences** using BERT-based and lexicon-based metrics.

	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%

Table 2: 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)

the advantage of being insensitive to the difference in score distributions across metrics.

Table 3 lists the results. They show that 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. 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.

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*	—	—

Table 3: 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 ( $p < 0.05$ )

## 5. Conclusions and Future Work

In this paper, we use a financial news dataset written in Romanian that are gathered between 2015 and 2019 from bursa.ro. We focused the research on a specific company, Banca Transilvania, and reduced the texts to paragraphs and even sentences containing the precise name. We have investigated the relationship between the sentiment expressed in news texts and the stock price for the bank, and showed there is a positive correlation between the polarity of the sentiment expressed and the price.

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 such as regression models for forecasting the future stock price based on the current portrayal in news texts, which could be a better indicator of a causality effect between the two.

## Acknowledgement

This work was supported by a grant of the Romanian Ministry of Education and Research, CC-CDI—UEFISCDI, project number 411PED/2020, code PN-III-P2-2.1-PED-2019-2271, within PNCDI III.

## References

- Agüero-Torales, M. M., Abreu Salas, J. I., and López-Herrera, A. G. (2021). Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing*, 107:107373.
- Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
- Barriere, V. and Balahur, A. (2020). Improving sentiment analysis over non-english tweets using multilingual transformers and automatic translation for data-augmentation. In *COLING*, pages 266–271, 01.
- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O’Reilly Media, Inc., 1st edition.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dumitrescu, S. D., Avram, A.-M., and Pyysalo, S. (2020). The birth of romanian bert. *arXiv preprint arXiv:2009.08712*.
- Gandhmal, D. P. and Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, 34:100190.
- Hutto, C. J. and Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Eytan Adar, et al., editors,

- Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*. The AAAI Press.
- Kaity, M. and Balakrishnan, V. (2019). An automatic non-english sentiment lexicon builder using unannotated corpus. *The Journal of Supercomputing*, 75:2243–2268.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kollintza-Kyriakoulia, F., Maragoudakis, M., and Krithara, A. (2018). Measuring the impact of financial news and social media on stock market modeling using time series mining techniques. *Algorithms*, 11(11).
- Li, X., Wu, P., and Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. *Information Processing & Management*, 57(5):102212.
- Lichtblau, D. and Stoean, C. (2018). Text documents encoding through images for authorship attribution. In Thierry Dutoit, et al., editors, *Statistical Language and Speech Processing*, pages 178–189, Cham. Springer International Publishing.
- Nemes, L. and Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, pages 1–20.
- Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., and Kim, H.-C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21).
- Shah, D., Isah, H., and Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, 7(2).
- Stoean, C. and Lichtblau, D. (2020). Author identification using chaos game representation and deep learning. *Mathematics*, 8(11).
- Stoean, C. and Lichtblau, D. (2021). Sentiment analysis from stock market news in romanian using chaos game representation. In *2021 23th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 252–258, December.
- Stoean, C., Paja, W., Stoean, R., and Sandita, A. (2019). Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations. *PLOS ONE*, 14(10):1–19, 10.
- Tache, A. M., Gaman, M., and Ionescu, R. T. (2021). Clustering word embeddings with self-organizing maps. application on laroseda—a large romanian sentiment data set. *arXiv preprint arXiv:2101.04197*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S., and Dong, X. (2021). Sentiment correlation in financial news networks and associated market movements. *Scientific reports*, (3062).
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al. (2019). Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.