Investigating the Relationship Between Romanian Financial News and Closing Prices of the Bucharest Stock Exchange

Ioan-Bogdan Iordache\textsuperscript{1,2}, Ana Sabina Uban\textsuperscript{1,2}, Catalin Stoian\textsuperscript{1,3}, Liviu P. Dinu\textsuperscript{1,2}

\textsuperscript{1}Human Language Technology Research Center, Bucharest, Romania, \textsuperscript{2}Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest, Romania, \textsuperscript{3}Faculty of Sciences, University of Craiova, Craiova, Romania

\textbf{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 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.

\textbf{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.

\textbf{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 \url{https://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 ($\approx 4.3\%$ 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.

\textbf{Experiments}

Since most available resources are developed for English, we include approaches where we first translate the Romanian texts into English and then use these resources. We consider both the translated text and the original text.

\textbf{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 stock price and the negative sentiment is anti-correlated with the stock price.

\begin{table}[h]
\centering
\begin{tabular}{lccc}
\hline
Metric & Correlation positive & Correlation negative & Correlation neutral \\
\hline
Vader Sentence-EN & 0.050* & 0.031 & 0.039 \\
Vader Paragraph-EN & 0.057* & -0.048 & -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* & -- & -- \\
\hline
\end{tabular}
\caption{Correlations between sentiment scores and stock price for different sentiment metrics at sentence and paragraph level, separately for the positive, negative, and neutral sentiment scores.}
\end{table}

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.

\textbf{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.

\textbf{References}