

## ABSTRACT

**mgr Piotr Łukasz Bednarski**

In today's financial markets, the speed of reaction to information is crucial for effective investment decision-making. This doctoral thesis proposes an innovative approach to forecasting stock prices using advanced Natural Language Processing (NLP) techniques and Long Short-Term Memory (LSTM) neural networks.

The main aim of this study was proposing a novel machine learning model architecture that utilizes NLP algorithms, including LSTM neural networks and BERT language models, for automatic processing and analysis of financial news content, and investigating their efficiency. In this context, traditional econometric models such as ARMA-GARCH served as a reference point.

The proposed model integrates two tools: BERT-like language models, which enable understanding the context and meaning of financial news, and LSTM networks, allowing for learning long-term dependencies in sequential data—specifically, time series financial data. To achieve this a novel FENE (Financial and Economic News Encoder) module was introduced with a goal of automatic news text analysis and extracting information, which may improve forecasts' quality.

Additionally, a hybrid approach was explored, combining traditional econometric models with the author's FENE module. This method leverages the advantages of both approaches, enhancing forecast accuracy by better utilizing available data and a more precise understanding of the context of financial information.

To evaluate the effectiveness of the proposed models, various forecast accuracy metrics based on forecast errors and statistical tests were applied.

The conducted research suggests that the author's approach based on LSTM neural networks and BERT-like language models may outperform traditional econometric methods, like ARMA-GARCH models, in forecasting stock prices based on financial news. Averaged RMSE error of the author's model is 4.56% lower than the same error for ARMA-GARCH models, and up to 25.4% lower for the most medial companies. Furthermore, hybrid models demonstrate the potential to improve forecast accuracy by leveraging synergies between different data analysis techniques. RMSE error of hybrid models is up to 1.92% lower than the error of classical econometric models. This work represents a significant step towards utilizing machine learning methods to better understand and predict the behaviour of stock market quotes.

The conclusions drawn from this work indicate a significant potential for utilizing textual data contained in financial news to improve stock price prediction processes. Integrating advanced NLP techniques with market forecasting models can contribute to more precise and effective investment decisions.