Data Science in Finance and Economics

DSFE, 5(1): 35–52.
DOI: 10.3934/DSFE.2025003
Received: 11 September 2024
Revised: 15 December 2024
Accepted: 21 January 2025
Published: 11 February 2025

https://www.aimspress.com/journal/dsfe

Research article

Are Natural Language Processing methods applicable to EPS forecasting in Poland?

Wojciech Kurylek*

Faculty of Management, University of Warsaw, 1/3 Szturmowa Street, 02-678 Warsaw, Poland

* **Correspondence:** Email: wkurylek@wz.uw.edu.pl.

Abstract: Accurate earnings forecasts are crucial for successful investment outcomes, especially in emerging markets like Poland, where analyst coverage is limited. This study investigated the applicability of natural language processing (NLP) techniques, specifically FastText and FinBERT word embeddings, combined with a gradient-boosting decision tree (XGBoost) machine learning algorithm, to forecast earnings per share (EPS) for companies listed on the Warsaw Stock Exchange from 2010 to 2019. The performance of these models was compared with a seasonal random walk (SRW) model. The SRW model consistently demonstrated the lowest error rates, as measured by the mean arctangent absolute percentage error, and outperformed the NLP-based models across different periods and error metrics. The superior performance of the simple SRW model can be attributed to the overparameterization and overfitting tendencies of the complex NLP models, as well as the relatively straightforward dynamics of the Polish stock market. The findings suggest that the application of sophisticated NLP techniques for EPS forecasting in Poland may not be justified, and that the SRW model provides a more accurate representation of the market's behavior.

Keywords: earnings per share; gradient-boosting decision tree; natural language processing; FastText; FinBERT; Warsaw Stock Exchange

JEL Codes: C01, C02, C12, C14, C58, G17

Abbreviations: MAAPE: mean arctangent absolute percentage error; NLP: natural language processing; SRW: seasonal random walk

1. Introduction

The valuation of company stocks critically depends on multiplying the earnings per share (EPS) by the price to earnings ratio, a fundamental aspect of investment analysis. Precise forecasting of these elements, particularly EPS, is crucial, as they provide key numerical insights into a company's future direction, offering valuable data on potential market valuation and guiding auditing expectations. While extensive financial analyst coverage is common for companies in developed markets, companies from emerging markets receive much less attention.

The Polish stock market, following its integration into the European Union after 2004, has demonstrated significant growth and development. As of the end of 2021, the market capitalization reached $197 billion, with 774 companies listed on the Warsaw Stock Exchange (WSE). However, despite this substantial market size, analyst coverage for Polish stocks remains limited compared with more developed markets like the United States or Western Europe. Approximately 20% of the 711 companies listed in 2019 received analyst attention, highlighting a significant gap in market coverage. This limited analyst coverage underscores the need for alternative methods, such as statistical or machine learning models, to forecast crucial financial metrics like EPS. Accurate EPS predictions are vital for investors, as they provide insights into a company's future performance and guide investment decisions. In the absence of extensive analyst coverage, the application of natural language processing (NLP) techniques to publicly available financial reports offer a promising avenue for improving EPS forecasting in the Polish market.

The Polish stock market's unique characteristics, including its size, structure, and limited analyst coverage, make it an ideal candidate for exploring the potential of NLP in EPS forecasting. The market's rapid growth and increasing importance within the European Union further emphasize the need for accurate and reliable forecasting methods. By leveraging the vast amount of textual data available in financial reports, NLP techniques can extract valuable insights and improve the accuracy of EPS predictions. Moreover, the successful application of NLP in EPS forecasting for the Polish market could have significant implications for other emerging markets with similar characteristics. Many developing countries face challenges related to limited analyst coverage, and the insights gained from this study could provide a framework for improving financial forecasting in these markets. Additionally, the findings of this research may also be relevant for smaller companies in developed markets that receive less analyst attention.

This article undertakes a comparative analysis of various models using a comprehensive set of explanatory variables derived from two different NLP text vectorization methods, namely FastText and FinBERT, applied to companies' public reports, employing the gradient-boosting decision tree (XGBoost) machine learning algorithm. In recent years, NLP techniques have been employed to derive new numerical features from textual documents in the fields of accounting and finance. These textual features have proven effective in predicting financial variables (Rawte et al., 2021; Medya et al., 2022), and models based on them can achieve substantial accuracy (Rao et al., 2022). However, existing research has not focused on EPS prediction using NLP tools but rather on predicting price movements. Closely related research by Ishikawa et al. (2020) applied NLP tools to forecast analysts' net income estimates, while Medya et al. (2022) demonstrated that the semantic features of financial texts were more predictive of stock price movements than traditional hard data, such as sales and EPS, in most cases. This study covers quarterly EPS data for 267 companies listed on the Polish stock exchange from the 2008–2009 financial crisis through to the 2020 pandemic.

Instead of solely using the conventional mean absolute percentage error, which can be distorted by small denominators, this study utilized an alternative measure, the mean arctangent absolute percentage error (MAAPE), as proposed by Kim and Kim (2016).

In summary, this article pursues multiple objectives: evaluating the performance of an advanced machine learning technique of gradient-boosted trees, using NLP-extracted features in EPS prediction for the Polish market, and applying diverse error metrics, varying timeframes, and statistical tests to validate the experimental outcomes. Additionally, it aimed to clarify the practical implications of these findings for investment strategies in Polish stocks.

2. Literature review

Algorithmic forecasting of EPS began in the 1960s, leading to scholarly investigation centered primarily on autoregressive integrated moving average (ARIMA) models (Ball and Watts, 1972; Watts, 1975; Griffin, 1977; Foster, 1977; Brown and Rozeff, 1979). These models were the primary focus, with the research findings showing variability: some studies advocated for the simplicity of the basic random walk model, suggesting that more complex models did not consistently outperform it, while others reported divergent conclusions. A similar examination of the Polish market was conducted by Kuryłek (2023a).

Over time, a preference emerged for ARIMA-type models due to their typically accurate forecasts (Lorek, 1979; Bathke and Lorek, 1984). This preference remained until the late 1980s, when a prevailing belief emerged that financial analysts' forecasts surpassed those generated by time series models (Brown et al., 1987). Conroy and Harris (1987) noted the superior performance of analysts' forecasts for short-term horizons, which diminished over longer periods. This view persisted until recent years, when the superiority of analysts over time series models was once again questioned (Lacina et al., 2011; Bradshaw et al., 2012; Pagach and Warr, 2020; Gatsios et al., 2021).

Since the late 1960s, various approaches employing exponential smoothing for EPS prediction have also been explored (Elton and Gruber, 1972; Ball and Watts, 1972; Johnson and Schmitt, 1974; Brooks and Buckmaster, 1976; Ruland, 1980; Brandon et al., 1987; Jarrett, 2008), producing mixed results. Research for the Polish market was undertaken by Kuryłek (2023b).

Lorek and Willinger (1996) demonstrated the superiority of multivariate cross-sectional models over firm-specific and common-structure ARIMA models. Lev and Thiagarajan (1993) identified 12 fundamental signals from financial ratios, subsequently utilized by Abarbanell and Bushee (1997) for EPS forecasting. Similar fundamental variables were employed by Cao and Gan (2009), Cao and Parry (2009), Etemadi et al. (2015), and Ball and Ghysels (2017) for multivariate EPS forecasting using neural networks, confirming their effectiveness. Ohlson (1995, 2001) formulated a residual earnings model, while Pope and Wang (2005, 2014) established theoretical frameworks linking earnings forecasts to accounting variables and stock prices. Li (2011) developed a model for forecasting earnings for loss-making firms, demonstrating its efficacy. Lev and Sougiannis (2010) provided evidence of the usefulness of estimate-based accounting items for predicting next year's earnings, albeit with limited success in subsequent years. Hou et al. (2012) achieved substantial R^2 coefficients in cross-sectional regression models for forecasting earnings. Li and Mohanram (2014) compared various models, revealing the superiority of some over others. Harris and Wang (2019) found Pope and Wang's (2005) model to be generally less biased and more accurate.

Decision tree-based techniques applied to EPS forecasting using accounting variables have yielded mixed results (Delen et al., 2013; Gerakos and Gramacy, 2013; Elamir, 2020; Chen et al., 2020). Recent advances in machine learning and deep learning have facilitated innovative experiments. Wang (2022) provided an overview of machine learning techniques for financial ratios forecasting, including EPS. Dreher et al. (2024) found that considering tax loss carryforwards did not improve EPS forecasts for German-listed companies and sometimes worsened predictions in out-of-sample tests, when using tax footnotes' information.

The domains of accounting, auditing, and finance frequently produce textual documents intended to communicate a wide array of messages, including but not limited to corporate financial performance, the management's assessment of current and future firm performance, domain standards, and regulations, as well as evidence of compliance with relevant standards and regulations. NLP applications have been used to mine these documents to obtain insights, make inferences, and create additional artifacts to advance knowledge in accounting, auditing, and finance. However, this is a relatively new and unexplored area of research, requiring expertise in both finance and computer engineering.

The paper by Fisher et al. (2016) synthesized the early literature on NLP in accounting, auditing, and finance to establish the state of knowledge at that time and to identify paths for future research. Since then, NLP methods have evolved substantially. Frankel et al. (2017) suggested that NLP techniques for narrative content from management conference calls and earnings prediction have reasonable construct validity, which is likely to be enhanced by further consideration of the unique characteristics of the text. More recently, Kambadura et al. (2023), in their book, presented core NLP techniques, mathematics in NLP, and applications.

A special area of research focuses on a specific class of NLP models. Xu (2019) used financial news from Reuters, Bloomberg, Yahoo Finance, etc., applying NLP to predict stock trends. Different word embedding techniques (i.e., vectorization of text) were compared. Huang et al. (2020) developed a financial text-specific variant of BERT (Bidirectional Encoder Representations from Transformers). In a Master's dissertation, Blomme and Dedeyne (2020) analyzed the impact of companies' financial reports on abnormal stock returns using FinBERT NLP methods. FinBERT, a customized version of BERT (a large language model by Google Inc.) for financial documents, was shown by Huang et al. (2023) to substantially outperform other NLP algorithms in sentiment classification for financial documents. FinBERT also excelled in identifying discussions related to environmental, social, and governance issues, outperforming other models by at least 18% in the textual informativeness of earnings conference calls.

Most existing research has focused on predicting stock prices using different NLP tools. Nabiee (2020) investigated the usefulness of changes in firms' annual and quarterly report language as signals for predicting future stock returns, using various similarity measures. Ishikawa et al. (2020) proposed and analyzed a methodology for forecasting the movements of analysts' net income estimates and stock prices using NLP and neural networks in the context of analyst reports. Differences were found among security firms depending on whether the analysts' net income estimates were forecasted by opinions or facts within reports, with trends in net income estimates proving effective for forecasting.

Wujec (2021) highlighted the increasing importance of text documents as a source of company information, using deep neural networks in NLP to analyze the sentiment of financial texts. The sentiment calculated using this method was directly related to market reactions. Rao et al. (2022) integrated public sentiment data into traditional stock analyses to predict stock price trends, using headlines from news publications and conversations from Yahoo! Finance's forum. The model achieved substantial accuracy. Medya et al. (2022) studied the statistical relationship among earnings calls, company sales, stock performance, and analysts' recommendations, with the semantic features of transcripts proving more predictive of stock price movements than traditional hard accounting data in most cases.

Santana Garcia (2023) analyzed the impact of financial news from two newspapers (the *Financial Times* and the *Wall Street Journal*) on the price dynamics of traded oil and gas companies, employing different NLP techniques, including fine-tuned large language models like BERT. Kiran et al. (2023) used sentiment analysis from various social media platforms to predict stock prices, employing different algorithms like K-means, long short-term memory, and convolutional neural networks for prediction, and error metrics like mean absolute error (MAE), mean standard error, and root mean standard error (RMSE).

Bv et al. (2023) analyzed the correlation between earnings call transcripts and subsequent stock price changes, extracting sentiments from quarterly transcripts using NLP techniques and identifying a possible causality between transcript sentiment and stock price movements. Rawte et al. (2021) emphasized the importance of studying similarities between documents over consecutive years in NLP for a comparative analysis of financial documents. Wang et al. (2024) examined the effect of competitive strategies on firm-level stock liquidity, using a machine-learning-based NLP approach, finding that firms with differentiation strategies showed higher stock liquidity due to increased investor attention and trading activities.

Recent studies have demonstrated the increasing application of NLP techniques in various aspects of finance, with a particular focus on the Polish market and Central and Eastern European (CEE) economies. These studies highlight the potential of NLP in enhancing financial forecasting, sentiment analysis, and content-based recommendations. Most of the studies have investigated the impact of sentiment on stock returns and trading volume in the Polish context. Polak (2021) examined the impact of sentiment derived from news headlines on the direction of stock price changes in the Polish stock market, finding no significant explanatory power in a one-day time frame. Wawer and Sobiczewska (2019) compared various NLP methods for predicting the sentiment of Polish-language short texts, finding strong performance from BERT and the multinomial Naive Bayes classifier. Wierzba et al. (2021) presented the Emotion Meanings dataset, a novel dataset of 6000 Polish word meanings derived from the Polish wordnet, annotated for valence, arousal, and basic emotion categories, providing a valuable resource for sentiment analysis in the Polish language. NLP techniques have also been applied to study the interdependencies between stock market indices and uncertainty measures in CEE economies. Kropiński (2023) investigated the relationship between main stock market indices and uncertainty measures quantified by X (formerly Twitter) messages for five CEE economies, including Poland.

The predictive power of narratives and economic research in the Polish market has also been explored using NLP techniques. Rybinski (2023) demonstrated that the content of news articles has a significant impact on their lifespan in five countries, including Poland. Rybinski (2021) presented a ranking of professional forecasters based on the predictive power of their research report narratives, showing that including NLP indexes in forecasting models lowers forecast errors in the Polish context. Rybinski (2020) compared the forecasting power of articles in a major Polish daily newspaper and regular research reports released by professional forecasters, finding that sell-side research may offer forecasting value, depending on the institution.

Dragan and Wróblewska (2019) adapted state-of-the-art NLP methods for determining similarities between text documents and content-based recommendations in the context of a Polish e-commerce platform, comparing various recommendation methods using semantic text analysis techniques to query the search engine index of documents. Klimczak (2020) presented the current state of the art in text analysis and its potential applications in finance, focusing on key innovations in sentiment analysis, topic identification, disambiguation, and multilingualism. Finally, the application of machine learning and NLP methods in literature reviews has been demonstrated by Łaniewski and Ślepaczuk (2024), who utilized these techniques to analyze and organize knowledge in the field of algorithmic trading, highlighting the effectiveness of large language models in refining datasets and addressing complex questions.

This study makes a notable contribution to the existing literature on EPS forecasting by investigating the applicability of NLP techniques, specifically FastText and FinBERT word embeddings, in the context of the Polish stock market. While previous research has explored various methods for EPS forecasting, including time series models, multivariate cross-sectional models, and machine learning techniques, the use of NLP-based approaches for this purpose has been relatively unexplored, particularly in emerging markets like Poland. The findings challenge the notion that sophisticated NLP techniques necessarily lead to improved EPS forecasting accuracy, as the SRW model consistently outperformed the NLP-based models across different periods and error metrics.

3. Data and methods

3.1. Data

Following its integration into the European Union after 2004, the Polish stock market demonstrated significant breadth, with market capitalization reaching $197 billion and encompassing 774 listed companies by the close of 2021. However, analyst coverage for these stocks remains limited relative to the United States or Western Europe, with approximately 20% of the 711 listed companies receiving attention in 2019. This limited coverage highlights the necessity for statistical or machine learning methods to forecast essential financial metrics. This research concentrates on the EPS series and other financial explanatory variables sourced from EquityRT, a financial analysis platform.

The EPS patterns of companies listed on the WSE are analyzed from the first quarter of 2010 to the fourth quarter of 2019, i.e. between notable structural changes such as the 2008–2009 financial crisis and the onset of the COVID-19 pandemic in 2020. For forecasting purposes, data from the first quarter of 2010 to the fourth quarter of 2018 (36 quarters) are used for model estimation, with data from the first quarter of 2019 to the fourth quarter of 2019 set aside for out-of-sample validation. Forecast horizons range from one to four quarters ahead, supplemented by additional validation samples from the years 2017 and 2018. The dataset, after ensuring comprehensive coverage and excluding any splits or reverse splits, includes 267 companies.

3.2. The Models

The seasonal random walk (SRW) model

Q_t denotes the realization of *EPS* in quarter *t*. The SRW model is described as:

$$Q_t = Q_{t-4} + \varepsilon_t \text{ where } \varepsilon_t \text{ are IID and } \varepsilon_t \sim N(0, \sigma^2) \tag{1}$$

The forecast given by $\widehat{EPS}_t = EPS_{t-4}$ employs the value from four quarters prior as the prediction, negating the necessity for parameter estimation. This model acts as a baseline, with its effectiveness in outperforming many time series models within the Polish context highlighted in the studies by Kuryłek (2023a, 2023b, 2024).

The multivariate models with word embeddings

The Elektroniczny System Przekazywania Informacji (ESPI) system functions as an electronic mechanism for public companies in Poland to disclose significant information. Utilized by entities listed on the WSE's main market, ESPI reports are obligatory to ensure prompt and transparent communication of crucial data such as financial results, major shareholder changes, corporate activities, and other pertinent events influencing stock prices or investor decisions. The oversight and dissemination of ESPI reports fall under the jurisdiction of the Polish Financial Supervision Authority (Komisja Nadzoru Finansowego), which enforces compliance among WSE-listed companies to uphold market transparency and safeguard investor interests.

Meanwhile, the Elektroniczna Baza Informacji (EBI) system serves as another digital reporting platform for Polish public companies. Predominantly used by companies on the NewConnect market (an alternative trading venue managed by the WSE for smaller and developing enterprises, particularly those in nascent stages or the tech sector), EBI reports are intended to ensure timely and transparent dissemination of essential information to investors and the public. While the WSE operates the EBI system, adherence to reporting mandates is ensured and supervised by the Komisja Nadzoru Finansowego, thus preserving market's integrity and protecting investors' interests.

Both ESPI and EBI systems leverage HTML5 (hypertext markup language version 5) to guarantee accessible, readable, and interactive reports across various devices and web browsers. Reports are provided by Notoria Serwis S.A. and categorized into 32 inherent types. For individual companies, text pre-processing involves the removal of HTML tags, URLs, email addresses, and nonalphanumeric characters, and conversion to lowercase, followed by extracting and merging the reports' topics and content. The sequentially ordered extracts for each company and quarter are combined into a string of words and numbers. This string is then transformed into vectors using word embeddings.

Word embeddings in NLP represent words as vectors of real numbers within a continuous vector space, allowing words with similar meanings to have related representations, effectively capturing semantic relationships and contexts. This technique employs dimensionality reduction to map words into a lower-dimensional space, facilitating efficient computations and contextually representing words based on their document usage, thus capturing semantic similarities. Foundational to various NLP tasks, word embeddings enhance models' text comprehension by embedding semantic meanings, significantly improving NLP applications' performance.

This study applies two types of word embeddings: FastText (developed by Meta Inc.) for the Polish language and FinBERT. FastText, an efficient open-source library for text classification and word representation, was introduced by Facebook AI Research (FAIR) with key contributions from Joulin et al. (2016). It achieves results comparable with those of deep learning models at reduced computational costs, making it suitable for large-scale text classification. A specific version trained by web crawling is applied to the Polish language. In contrast, FinBERT (Huang et al., 2020) is a sophisticated NLP model designed specifically for financial text embeddings. Although FinBERT is not language-specific, it applies to various languages, including Polish. As a domain-specific variant of BERT (Bidirectional Encoder Representations from Transformers), developed by Google Inc.

(Chang et al., 2018), it utilizes the Transformer architecture, a deep learning paradigm that has revolutionized NLP. Transformers, with their attention mechanism, focus on relevant sentence parts during processing, effectively capturing long-range word dependencies. They consist of an encoder and decoder: the encoder creates contextual word representations, and the decoder generates the desired output by focusing on the pertinent encoded information.

To represent an entire document as a vector, word vectors must be combined into a single vector for the document. Document embeddings, which condense whole documents or text sequences into dense vectors, can be constructed using pooling operations such as a simple mean, as employed in this study. Consequently, the multivariate model utilizing word embeddings can be articulated with FastText and FinBERT embeddings in the following way:

$$EPS_{t+4,i} = f\left(EPS_{t,i}, N_{t,i}^{1}, N_{t,i}^{2}, .., N_{t,i}^{n}, X_{t,i}^{1}, X_{t,i}^{2}, .., X_{t,i}^{m}\right) + \varepsilon_{t,i} \tag{2}$$

where $N_{t,i}^{j}$ is the number of reports in the *j*-th category for the *i*-th company in the *t*-th quarter within *n* categories, and *n* equals 32. The vector $\left[X_{t,i}^{1}, X_{t,i}^{2}, .., X_{t,i}^{m}\right]$ represents document embeddings of reports for *i*-th company in the *t*-th quarter. For the FastText word embeddings, the vector length (denoted as the parameter *m*) is 300 dimensions, whereas for FinBERT embeddings, it is 512 dimensions. Thus, incorporating the delayed EPS variable and 32 categorical variables results in independent variable vectors of 333 and 545 dimensions for FastText and FinBERT, respectively. The model was trained on a dataset comprising 7476 observations (calculated from 28 quarters across 267 companies) to forecast EPS for 2019. Implementation of the FastText model was carried out using the "fasttext" library in Python, while the "transformers" library facilitated the FinBERT model.

3.3. Estimation techniques – a gradient-boosting decision tree (XGBoost)

Initially introduced by Chen and Guestrin in 2016, XGBoost (eXtreme Gradient Boosting) represents a substantial enhancement over the traditional gradient boosting algorithm, known for its notable speed and effectiveness. This advanced machine learning approach, which traces its roots to the decision tree methodologies developed in the 1960s, is extensively utilized for both regression and classification tasks. It develops a predictive model by combining numerous weak learners, usually simple decision trees, in an ensemble manner. Each successive tree aims to correct the mistakes of its predecessors, with the gradient descent algorithm iteratively updating the weights of these weak learners. This process continues until the loss function is minimized or a specific stopping criterion is achieved. XGBoost incorporates several techniques to improve the performance of gradient boosting models, including regularization to reduce overfitting by adding penalties to the loss function, tree pruning to eliminate unnecessary branches and enhance the model's stability, and parallelization to accelerate the training process. Additionally, XGBoost effectively captures nonlinearities in the data. For a comprehensive overview, Simon's 2020 book provides valuable insights. The implementation of this methodology utilizes the "xgb" library in Python, with the hyperparameters optimized (by cross-validation in the training sample) using the "hyperopt" library to enhance forecast accuracy (Banerjee, 2020).

It has been documented by some researchers (Armon et al., 2022; Borisov et al., 2024) that gradient-boosting models such as XGBoost can frequently be effectively used in tabular datasets, particularly with structured and relational data, due to their ability to naturally incorporate features' interactions.

3.4. Mean arctangent absolute percentage error

The actual EPS for a particular company i across the first to the fourth quarters of 2019 can be denoted by $A_1^i, \ldots, A_4^i$. The corresponding forecasts for these quarters, represented as $F_1^i, \ldots, F_4^i$ (i.e. $\hat{Q}_t$, where t = 37,..,40 for the i-th company), allow the absolute percentage error (APE) for the forecast in the j-th quarter of 2019 for any i-th firm to be expressed as follows:

$$APE_j^i = \left|\frac{A_j^i - F_j^i}{A_j^i}\right| \tag{3}$$

The APE, despite its widespread use, has a critical limitation: it tends to produce infinite or undefined values when the actual figures are near zero, a frequent occurrence in earnings predictions. Furthermore, extremely low actual values, typically less than one, can result in disproportionately high percentage errors, causing outliers. This issue is exacerbated when actual values are zero, leading to infinite or undefined APEs. To address this problem, the arctangent APE was introduced by Kim and Kim (2016), offering a more robust alternative for handling such cases in predictive analytics.

$$AAPE_j^i = \begin{cases} 0 & if\ A_j^i = F_j^i = 0 \\ arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) & otherwise \end{cases} \tag{4}$$

The basis for adopting this method stems from the characteristic of the arctangent function, which maps values ranging from negative infinity to positive infinity into the interval [-π/2, π/2]. As a result, the MAAPE for the j-th quarter across all I companies in the dataset can be expressed as follows:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE_j^i = \frac{1}{I}\sum_{i=1}^{I} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{5}$$

The preference for MAAPE instead of mean APE was intentional because of the occurrence of companies with actual profits approaching zero in the dataset under study. In cases where only one observation approaches zero while the others are significantly distant, the mean APE for that particular observation can escalate to an extremely high value, approaching infinity. This issue has the potential to dominate the mean calculation, making the remaining observations negligible in their impact.

3.5. The statistical test

To assess the statistical significance of differences in the MAAPE across multiple models, a nonparametric two-sided Wilcoxon test, originally described by Wilcoxon (1945), was employed. This test serves as a paired difference test for two matched samples. It is important to note that this test does not require specific assumptions about the probability distribution, except for the symmetry of score differences and the independence of these differences. Ruland (1980) extensively discussed the application of the Wilcoxon test in validation, particularly in determining whether errors from various EPS models exhibit statistically significant distinctions. Separate tables containing p-values were generated for each quarter, ranging from one to four, as well as for all quarters combined.

$$H_o: AAPEs\ of\ a\ pair\ of\ models\ are\ the\ same \tag{6}$$

If the *p*-values from each test fall below the predefined significance level of 0.05, the null hypothesis for each test will be rejected. This widely accepted principle is supported by various sources, including Ruland (1980).

4. Results

The SRW model, detailed in Table 1, consistently demonstrated superior performance over all multivariate word embedding models across each quarter and throughout 2019, exhibiting the lowest MAAPE overall. In contrast, the model utilizing FinBERT word embeddings showed the poorest performance among all methods, with the FastText approach falling between these two extremes.

To determine whether the errors of the top-performing model significantly differ from those of the other methods, the Wilcoxon nonparametric test was utilized to compare the arctangent APE medians between the SRW model and all alternative approaches. As shown in Table 2, the results indicate that the SRW model consistently exhibits statistically lower errors compared with the other methods across all analyzed periods of 2019.

Table 1. Summary statistics on forecast errors and mean equality tests for the quarters of 2019.

		Quarters		All quarters	
Model	Q1 MAAPE	Q2 MAAPE	Q3 MAAPE	Q4 MAAPE	MAAPE
SRW	0.658	0.702	0.653	0.736	0.687
FastText	0.896	0.800	0.914	0.892	0.876
FinBERT	1.093	1.068	1.060	1.044	1.066

Table 2. *P*-values of the Wilcoxon test of forecast errors for SRW and respective models in 2019.

Quarter	Model	FastText	FinBERT
1	SRW	0.000	0.000
2	SRW	0.000	0.000
3	SRW	0.000	0.000
4	SRW	0.000	0.000
All	SRW	0.000	0.000

4.1. Robustness checks

Robustness assessments were conducted across multiple years using two widely recognized error metrics. It is important to note that across the years examined (2017, 2018, and 2019) the SRW model consistently delivered superior forecasts compared with the other methods, as depicted in Table 3. However, the performance of the FinBERT approach varied; while it underperformed in 2019 and 2017, it marginally outperformed FastText in 2018.

Table 3. Summary statistics on forecast errors and mean equality tests for all quarters, 2017–2019.

Model	2017 MAAPE	2018 MAAPE	2019 MAAPE
SRW	0.686	0.711	0.687
FastText	0.810	0.875	0.876
FinBERT	0.888	0.843	1.066

Furthermore, the Wilcoxon test was employed to compare each model pair against the SRW model, with the corresponding *p*-values for each year detailed in Table 4. Throughout these years, the SRW model consistently exhibited statistically superior results compared with alternative methods. Hence, the sustained dominance of the SRW model over time is evident.

Table 4. *P*-values of the paired Wilcoxon test of forecast errors for all quarters of 2017–2019 and the SRW model.

Year	Model	FastText	FinBERT
2017	SRW	0.000	0.000
2018	SRW	0.000	0.000
2019	SRW	0.000	0.000

Table 5 provides a performance assessment of the analyzed models using alternative error measures: RMSE and MAE. This evaluation aggregates all quarters for the year 2019. To ensure an equitable comparison, these metrics were adjusted for consumer price index (CPI) inflation, ensuring that future errors' present values in nominal terms are equivalent to the current errors.

Consistent with previous findings in 2019, the SRW model demonstrated the lowest errors across all metrics, encompassing both RMSE and MAE. In contrast, the model utilizing FinBERT word embeddings exhibited the poorest performance in 2019 according to both error criteria, aligning with previous observations regarding MAAPE. Table 6 presents the *p*-values derived from the Wilcoxon test, highlighting significant differences between the outcomes of the SRW model and other model comparisons.

Table 5. RMSE and MAE in all joint quarters of 2019.

	SRW	FastText	FinBERT
RMSE	0.937	1.751	2.477
MAE	0.705	1.280	1.877

Table 6. *P*-values of the paired Wilcoxon test of forecast errors for RMSE and MAE in 2019.

Measure	Model	FastText	FinBERT
RMSE	SRW	0.000	0.000
MAE	SRW	0.000	0.000

The research findings suggest that despite incorporating variables from various word embedding algorithms and employing advanced forecasting techniques based on them, there was no observed improvement in performance. Moreover, these sophisticated approaches did not surpass the basic SRW method in terms of predictive accuracy.

5. Discussion

The comparatively lower performance of word embedding models, utilizing gradient-boosting decision trees, stems from issues related to overfitting. Overfitting causes unstable relationships among variables, which may hold only for specific test datasets, and can occur when the model becomes too complex and starts capturing noise in the training data rather than generalizable patterns. One of the best techniques to reduce the overfitting problem in the XGBoost model is the optimization of hyperparameters with cross-validation in the training sample, and this approach was applied. However, regardless of the use of this method, overfitting can persist. Using such relationships for predictions is reasonable only if their statistical robustness is sufficiently validated, as emphasized by Lev and Souginannis (2010). This observation is consistent with Dreher et al.'s (2024) findings for German listed firms, showing that while complex deep learning techniques enhance explanatory power within the sample, they fail in out-of-sample predictions. Such sophisticated models risk overparameterizing straightforward market behaviors, resulting in larger forecast errors. Additionally, introducing more variables than simply the EPS value delayed by four quarters adds unnecessary noise to the data, exacerbating the challenge of forecasting.

The rationale behind the superior performance of simpler models may resonate within the Polish context. Advanced models often become overly intricate, employing an excess of parameters to describe relatively straightforward economic phenomena. This observation echoes the findings of Kuryłek (2023a, 2023b, 2024), who demonstrated that basic models such as ARIMA and exponential smoothing, and also advanced time series methods, which are effective in the US market, were surpassed by the simple SRW model in Poland. This underscores the hypothesis that the inherent simplicity of the Polish stock market likely underpins the effectiveness of the SRW model, suggesting that additional calibration for out-of-sample predictions may be necessary.

Thus, the straightforward application of sophisticated techniques that are reliant on word embeddings beyond the conventional SRW for EPS forecasting in Polish investment contexts appears to be impractical. Moreover, considering that EPS behavior follows a SRW and acknowledging that stock prices derive from EPS multiplied by the P/E multiple, one might infer that stock prices exhibit at least as much randomness as EPS. Given this EPS behavior characterized by a random walk, accurately predicting stock prices even just one quarter ahead becomes particularly daunting.

In shorter timeframes where the EPS remains constant, stock price forecasting behaves similarly to P/E multiples. Therefore, exploring methods to forecast P/E multiples for periods shorter than one quarter, occurring between the publication of quarterly financial reports, could prove highly relevant from an investment standpoint. The forecast generated by the SRW essentially reflects a value from the corresponding quarter of the previous year. This implies that for predicting future prices, even over extended horizons, the P/E multiple might carry more weight than next year's earnings of companies (EPS). This conclusion aligns with economic theory, positing that the P/E multiple is influenced by anticipated growth in future earnings, future interest rates, and market sentiment or premium reflecting investors' risk appetite, while EPS forecasts relate solely to imminent earnings. In both short-term and long-term investment contexts, the consensus is clear: the P/E multiple surpasses EPS prediction in significance.

6. Conclusions

The research offers a significant contribution to the literature on EPS forecasting by examining the potential of NLP methods within the Polish stock market context. While past studies have extensively explored EPS forecasting through various approaches such as time series analysis, multivariate cross-sectional models, and machine learning algorithms, the integration of NLP-based techniques remains underexplored, especially in emerging markets like Poland.

The study examined the predictive capabilities of three methodologies, namely the SRW and a gradient boosting tree (XGBoost) utilizing two different word embedding models (FastText and FinBERT), applied to publicly available company reports. These multivariate approaches were trained using 333 and 545 explanatory variables, respectively, encompassing firm-specific textual attributes. EPS forecasting is crucial in emerging markets like Poland, where financial analyst coverage of listed companies is sparse. Analyzing quarterly EPS data from 267 Polish firms spanning 2010 to 2019, the SRW model consistently demonstrated the lowest error rates, providing a more accurate representation of the Polish market compared with alternative models. Additionally, the SRW model consistently outperformed other methods across different periods and error metrics such as RMSE or MAE. This pattern is supported by Wilcoxon tests and can be attributed to the overparameterization of complex models, their susceptibility to overfitting, and the relatively straightforward nature of the Polish stock market.

The research findings indicate that despite incorporating variables derived from word embeddings and employing advanced forecasting algorithms such as XGBoost on tabular data, there was no improvement in performance, challenging the assumption that advanced NLP methods inherently enhance EPS forecasting accuracy. Consequently, the practical implication of this research suggests that employing techniques beyond the standard SRW for EPS forecasting in Poland lacks justification. However, relying on the SRW for EPS modeling implies that forecasted stock prices may exhibit significant randomness, posing challenges for accurate prediction. Therefore, forecasting the P/E multiple might hold greater significance than predicting EPS for future stock price projections, especially in shorter investment horizons where the EPS remains stable.

Future research could explore the relationship between forecasting accuracy and firm size, with industry sector analysis potentially influencing the choice of the most suitable EPS forecasting model. Investigating time series transformations to normalize EPS distributions could offer valuable insights. Furthermore, evaluating the performance and accuracy of various predictive models and financial analysts' forecasts during economic downturns, such as the 2008–2009 financial crisis or the COVID-19 pandemic, could provide valuable perspectives. One could also explore the impact of different text sources on the NLP model's performance. While the current study focuses on public company reports, incorporating additional sources such as news articles, social media posts, or earnings call transcripts may provide a richer set of textual features and potentially improve the model's predictive power. Future researchers can investigate to find out which cases will yield better results than the SRW model when using the NLP-based model. Identifying seasonal patterns through the SRW model may also contribute insights into investment strategies, potentially challenging the "weak form" of the efficient market hypothesis.

Use of AI tools declaration

The author affirms that no artificial intelligence (AI) tools have been used in the creation of this work.

Funding

The authors have received no funding from any source in the preparation of this work.

Conflict of interest

The author declares no conflicts of interest in this paper.

References

Abarbanell J, Bushee B (1997) Fundamental analysis, future EPS, and stock prices. *J Account Res* 35: 1–24. https://doi.org/10.2307/2491464

Armon A, Shwartz-Ziv R (2022) Tabular data: Deep learning is not all you need. *Inform Fusion* 81: 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Ball R, Ghysels E (2017) Automated earnings forecasts: Beat analysts or combine and conquer? *Manage Sci* 64: 4936–4952. https://doi.org/10.1287/mnsc.2017.2864

Ball R, Watts R (1972) Some Time Series Properties of Accounting Income. *J Financ* 27: 663–681. http://dx.doi.org/10.1111/j.1540-6261.1972.tb00991.x

Banerjee P (2020) A Guide on XGBoost hyperparameters tuning, Accessed June 14, 2024. Available from: https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning.

Bathke Jr AW, Lorek KS (1984) The Relationship between Time-Series Models and the Security Market's Expectation of Quarterly Earnings. *Account Rev* 59: 163–176.

Blomme S, Dedeyne J (2020) Predicting the effect of 10-K, 10-Q and 8-K company reports on abnormal stock returns using FinBERT NLP methods. Master thesis in Business Engineering: Data Analytics, Faculteti Economie en Bedrufskunde. University of Gent.

Bv N, Simha JB, Abhi S (2023) Deploying NLP Techniques for Earnings Call Transcripts for Financial Analysis: A Reverse Phenomenon Paradigm. *7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2023 – Proceedings*: 368–375. https://doi.org/10.1109/I-SMAC58438.2023.10290494

Borisov V, Haug J, Kasneci G, et al. (2024) Deep Neural Networks and Tabular Data: A Survey. *Ieee T Neur Net Lear* 35: 7499–7519. https://doi.org/10.1109/tnnls.2022.3229161

Bradshaw M, Drake M, Myers J, et al. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Rev Account Stud* 17: 944–968. http://dx.doi.org/10.1007/s11142-012-9185-8

Brandon Ch, Jarrett JE, Khumawala SB, et al. (1987) A Comparative Study of the Forecasting Accuracy of Holt - Winters and Economic Indicator Models of Earnings Per Share for Financial Decision Making. *Manage Financ* 13: 10–15. http://dx.doi.org/10.1108/eb013581

Brooks LD, Buckmaster DA (1976) Further Evidence of The Time Series Properties of Accounting Income. *J Financ* 31: 1359–1373. http://dx.doi.org/10.1111/j.1540-6261.1976.tb03218.x

Brown LD, Griffin PA, Hagerman RL, et al. (1987) Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *J Account Econ* 9: 61–87. http://dx.doi.org/10.1016/0165-4101(87)90017-6

Brown LD, Rozeff MS (1979) Univariate Time-Series Models of Quarterly Accounting Earnings per Share: A Proposed Model. *J Account Res* 17: 179–189. http://dx.doi.org/10.2307/2490312

Cao Q, Gan Q (2009) Forecasting EPS of Chinese listed companies using a neural network with genetic algorithm. *15th Americas Conference on Information Systems* 2009*, AMCIS* 2009: 2791–2981.

Cao Q, Parry M (2009) Neural network earnings per share forecasting models: A comparison of backward propagation and the genetic algorithm. *Decis Support Syst* 47: 32–41. http://dx.doi.org/10.1016/j.dss.2008.12.011

Chang MW, Devlin J, Lee K, et al. (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*. arXiv, arXiv:1810.04805.

Chen Y, Chen S, Huang H, et al. (2020) Applied identification of industry data science using an advanced multi-componential discretization model. *Symmetry* 12: 1–28. https://doi.org/10.3390/sym12101620

Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Conroy R, Harris R (1987) Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods. *Manage Sci* 33: 725–738. http://dx.doi.org/10.1287/mnsc.33.6.725

Delen D, Kuzey C, Uyar A, et al. (2013) Measuring firm performance using financial ratios: A decision tree approach. *Expert Syst Appl* 40: 3970–3983. https://doi.org/10.1016/j.eswa.2013.01.012

Dragan Ł, Wróblewska A (2019) Content-Based Recommendations in an E-Commerce Platform. *Information Technology, Systems Research, and Computational Physics*, Springer International Publishing, 252–263. https://doi.org/10.1007/978-3-030-18058-4_20

Dreher S, Eichfelder S, Noth F, et al. (2024) Does IFRS information on tax loss carryforwards and negative performance improve predictions of earnings and cash flows? *J Bus Econ* 94: 1–39. http://dx.doi.org/10.1007/s11573-023-01147-7

Elamir E (2020) Modeling and predicting earnings per share via regression tree approaches in banking sector: Middle East and North African countries case. *Invest Manag Financ Innov* 17: 51–68. https://doi.org/10.21511/imfi.17(2).2020.05

Elton EJ, Gruber MJ (1972) Earnings Estimates and the Accuracy of Expectational Data. *Manage Sci* 18: B409–B424. http://dx.doi.org/10.1287/mnsc.18.8.B409

Etemadi H, Ahmadpour A, Moshashaei S, et al. (2015) Earnings Per Share Forecast Using Extracted Rules from Trained Neural Network by Genetic Algorithm. *Computational Econ* 46: 55–63. https://doi.org/10.1007/s10614-014-9455-6

Fisher IE, Garnsey MR, Hughes ME, et. al (2016) Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intell Syst Account* 23: 157–214. Portico. https://doi.org/10.1002/isaf.1386

Foster G (1977) Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results. *Account Rev* 52: 1–21.

Frankel RM, Jennings JN, Lee JA, et al. (2017) Using Natural Language Processing to Assess Text Usefulness to Readers: The Case of Conference Calls and Earnings Prediction. *SSRN Electronic J*. https://doi.org/10.2139/ssrn.3095754

Gatsios RC, Lima FG, Gaio LE, et al. (2021) Re-examining analyst superiority in forecasting results of publicly-traded Brazilian companies. *Revista de Administracao Mackenzie* 22: eRAMF210164. https://doi.org/10.1590/1678-6971/eramf210164

Gerakos J, Gramacy R (2013) Regression-Based Earnings Forecasts. *Chicago Booth Res Paper,* 12–26. https://doi.org/10.2139/ssrn.2112137

Griffin P (1977) The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence. *J Accounting Res* 15: 71–83. http://dx.doi.org/10.2307/2490556

Harris RDF, Wang P (2019) Model-based earnings forecasts vs. financial analysts' earnings forecasts. *British Account Rev* 51: 424–437. https://doi.org/10.1016/j.bar.2018.10.002

Hou K, van Dijk M, Zhang Y, et al. (2012) The implied cost of capital: A new approach. *J Account Econ* 53: 504–526. https://doi.org/10.1016/j.jacceco.2011.12.001

Huang AH, Wang H., Yang Y, et al. (2023) FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemp Account Res* 40: 806–841. Portico. https://doi.org/10.1111/1911-3846.12832

Huang D, Huang K, Liu Z, et al. (2020) FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence.* https://doi.org/10.24963/ijcai.2020/622

Ishikawa Y, Izumi K, Matsushima H, et al. (2020) Forecasting Net Income Estimate and Stock Price Using Text Mining from Economic Reports. *Information* 11: 1–21. https://doi.org/10.3390/info11060292

Jarrett JE (2008) Evaluating Methods for Forecasting Earnings Per Share. *Manage Financ* 16: 30–35. http://dx.doi.org/10.1108/eb013647

Johnson TE, Schmitt TG (1974) Effectiveness of Earnings Per Share Forecasts. *Financ Manage* 3: 64–72. http://dx.doi.org/10.2307/3665292

Joulin A, Grave E, Bojanowski P, et al. (2016) Bag of Tricks for Efficient Text Classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2, Short Papers.* https://doi.org/10.18653/v1/e17-2068

Kambadura P, Manna G, Stentb A, et al. (2023) NLP in Finance, In: Capponi A and Lehalle Ch A, *Machine Learning and Data Sciences for Financial Markets*, Cambridge University Press, Cambridge. https://doi.org/10.1080/14697688.2023.2280101

Kim S, Kim H (2016) A new metric of absolute percentage error for intermittent demand forecasts. *Int J Forecast* 32: 669–679. http://dx.doi.org/10.1016/j.ijforecast.2015.12.003

Kiran JS, Jonnalagadda S, Naga Veera Tarun D, et al. (2023) Stock Market Prediction Using Sentiment Analysis and Incremental Clustering Approaches. *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS* 2023: 888–893. https://doi.org/10.1109/ICACCS57279.2023.10112768

Klimczak KM (2020) Text analysis in finance: The challenges for efficient application. In: Gąsiorkiewicz, L., & Monkiewicz, J. (Eds.) *Innovation in Financial Services*, 199–216. https://doi.org/10.4324/9781003051664-4

Kropiński P (2023). Investigating Whether Economic Policy Uncertainty Affects Central and Eastern European Markets. Evidence from Twitter-Based Uncertainty Measures. *Available at SSRN* 4359895. https://doi.org/10.2139/ssrn.4359895

Kuryłek W (2023a) The modeling of earnings per share of Polish companies for the post-financial crisis period using random walk and ARIMA models. *J Bank Financ Econ* 1: 26–43. http://dx.doi.org/10.7172/2353-6845.jbfe.2023.1.2

Kuryłek W (2023b) Can exponential smoothing do better than seasonal random walk for earnings per share forecasting in Poland? *Bank Credit* 54: 651–672.

Kurylek W (2024) Can we profit from BigTechs' time series models in predicting earnings per share? Evidence from Poland. *Data Sci Financ Econ* 4: 218–235. http://dx.doi.org/10.3934/DSFE.2024008

Lacina M, Lee B, Xu R, et al. (2011) An evaluation of financial analysts and naïve methods in forecasting long-term earnings. In: Lawrence K D, and Klimberg R K (Eds.), *Advances in business and management forecasting*, Bingley, UK, Emerald, 77–101. http://dx.doi.org/10.1108/S1477-4070(2011)0000008009

Lev B, Souginannis T (2010) The usefulness of accounting estimates for predicting cash flows and earnings. *Rev Account Stud* 15: 779–807. http://dx.doi.org/10.1007/s11142-009-9107-6

Lev B, Thiagarajan S (1993) Fundamental information analysis. *J Account Res* 31: 190–215. http://doi.org/10.2307/2491270

Li KK (2011) How well do investors understand loss persistence? *Rev Account Stud* 16: 630–667. https://doi.org/10.1007/s11142-011-9157-4

Li KK, Mohanram P (2014) Evaluating cross-sectional forecasting models for the implied cost of capital. *Rev Account Stud* 19: 1152–1185. https://doi.org/10.1007/s11142-014-9282-y

Lorek KS (1979) Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models. *J Account Res* 17: 190–204. http://dx.doi.org/10.2307/2490313

Lorek KS, Willinger GL (1996) A multivariate time-series model for cash-flow data. *Accoun Rev* 71: 81–101.

Łaniewski S, Ślepaczuk R (2024). Enhancing literature review with NLP methods Algorithmic investment strategies case. *Faculty of Economic Studies, University of Warsaw Working Papers*. https://doi.org/10.33138/2957-0506.2024.16.452

Medya S, Rasoolinejad M, Uzzi B, et. al (2022) An Exploratory Study of Stock Price Movements from Earnings Calls. *WWW 2022 - Companion Proceedings of the Web Conference* 2022: 20–31. https://doi.org/10.1145/3487553.3524205

Nabiee S (2020) Prediction of Firms' Annual and Quarterly Return Using NLP Techniques. Master thesis in Electrical Engineering, University of California, Irvine.

Ohlson JA (1995) Earnings, Book Values, and Dividends in Equity Valuation. *Contemp Account Res* 11: 661–687. https://doi.org/10.1092/7tpj-rxqn-tqc7-ffae

Ohlson JA (2001) Earnings, Book Values, and Dividends in Equity Valuation: An Empirical Perspective. *Contemp Account Res* 18: 107–120. https://doi.org/10.1092/7tpj-rxqn-tqc7-ffae

Pagach DP, Warr RS (2020) Analysts versus time-series forecasts of quarterly earnings: A maintained hypothesis revisited. *Adv Account* 51: 1–15. http://dx.doi.org/10.1016/j.adiac.2020.100497

Polak K (2021) The Impact of Investor Sentiment on Direction of Stock Price Changes: Evidence from the Polish Stock Market. *J Bank Financ Econ* 2: 72–90. https://doi.org/10.7172/2353-6845.jbfe.2021.2.4

Pope PF, Wang P (2005) Earnings Components, Accounting Bias and Equity Valuation. *Rev Account Stud* 10: 387–407. https://doi.org/10.1007/s11142-005-4207-4

Pope P, Wang P (2014) On the relevance of earnings components: Valuation and forecasting links. *Rev Quant Financ Account* 42: 399–413. https://doi.org/10.1007/s11156-013-0347-y

Rao J, Ramaraju V, Smith J, et al. (2022) A Sentiment Analysis Based Stock Recommendation System. *Proceedings - 2022 IEEE 5th International Conference on Artificial Intelligence and Knowledge Engineering, AIKE* 2022: 82–89. https://doi.org/10.1109/AIKE55402.2022.00020

Rawte V, Gupta A, Zaki MJ, et al. (2021) A Comparative Analysis of Temporal Long Text Similarity: Application to Financial Documents. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 77–91. https://doi.org/10.1007/978-3-030-66981-2_7

Ruland W (1980) On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings. *Financ Manage* 9: 30–37. http://dx.doi.org/10.2307/3665165

Rybinski K (2020) Should asset managers pay for economic research? A machine learning evaluation. *J Financ Data Sci* 6: 31–48. https://doi.org/10.1016/j.jfds.2020.08.001

Rybinski K (2021) Ranking professional forecasters by the predictive power of their narratives. *Int J Forecast* 37: 186–204. https://doi.org/10.1016/j.ijforecast.2020.04.003

Rybinski K (2023) Content still matters. A machine learning model for predicting news longevity from textual and context features. *Inf Process Manage* 60: 103398. https://doi.org/10.1016/j.ipm.2023.103398

Santana García F (2023) The effect of financial news on stock prices: insights from NLP techniques. *Comillas Pontifical University, Faculty of Economics and Business Administration, ICADE Working Paper*.

Simon J (2020) Learn Amazon SageMaker: A guide to building, training, and deploying machine learning models for developers and data scientists, Packt> Birmingham – Mumbai.

Wang X, Han R, Zheng M et al. (2024) Competitive strategy and stock market liquidity: a natural language processing approach. *Inf Technol Manage* 25: 99–112. https://doi.org/10.1007/s10799-023-00401-2

Wawer A, Sobiczewska J (2019) Predicting Sentiment of Polish Language Short Texts. Proceedings - Natural Language Processing in a Deep Learning World, 1321–1327. https://doi.org/10.26615/978-954-452-056-4_151

Watts RL (1975) The Time Series Behavior of Quarterly Earnings. *Working paper, Department of Commerce,* University of New Castle, April 1975.

Wierzba M, Riegel M, Kocoń J, et al. (2021) Emotion norms for 6000 Polish word meanings with a direct mapping to the Polish wordnet. *Behav Res Methods* 54: 2146–2161. https://doi.org/10.3758/s13428-021-01697-0

Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics* 1: 80–83. http://dx.doi.org/10.2307/3001968

Wujec M (2021) Analysis of the Financial Information Contained in the Texts of Current Reports: A Deep Learning Approach. *J Risk Financ Manage* 14: 582. https://doi.org/10.3390/jrfm14120582

Wang XQ (2022) Research on enterprise financial performance evaluation method based on data mining. In: *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*. https://doi.org/10.1109/icetci55101.2022.9832404

Xu Z (2019) NLP driven large scale financial data analysis. Doctoral dissertation, University of Illinois at Urbana-Champaign.