

Article

Information Environment Quantifiers as Investment Analysis Basis

Dmitry G. Rodionov ¹, Polina A. Pashinina ¹, Evgenii A. Konnikov ^{1,*} and Olga A. Konnikova ²

¹ Graduate School of Industrial Economics, Peter the Great St. Petersburg Polytechnic University, 195251 Saint Petersburg, Russia

² Department of Marketing, Saint-Petersburg State University of Economics, 191023 Saint Petersburg, Russia

* Correspondence: konnikov.evgeniy@gmail.com; Tel.: +7-9618084582

Abstract: The combination of the processes of widespread digitalization and globalization of the world economy has led to a significant expansion of the global information environment. The modern information environment is dynamically active, and changes in it are indicators of changes in the material world. This specificity can be used for investment analysis purposes. However, at the time of this research, a universal methodology for analyzing the information environment has not yet been formed. The purpose of this study is to develop tools for quantifying the information environment and testing them as investment predictors. The key result of this study is a stock price forecasting model based on information environment quantifiers and its critical analysis. The results obtained will be useful both for investors of different skill levels and for researchers of the information environment.

Keywords: information environment; quantification; financial market; tokenization; tonality



Citation: Rodionov, Dmitry G., Polina A. Pashinina, Evgenii A. Konnikov, and Olga A. Konnikova. 2022. Information Environment Quantifiers as Investment Analysis Basis. *Economies* 10: 232. <https://doi.org/10.3390/economies10100232>

Academic Editor: Nick Drydakis

Received: 22 July 2022

Accepted: 5 September 2022

Published: 20 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The number of Russian investors is expected to multiply soon. The reason for this assumption is that due to the economic downturn in 2020 caused by the COVID-19 pandemic, the Central Bank reduced its key rate, which also means lower interest on bank deposits (in 2020). In addition, bank deposit tax was introduced on 1 January 2021, which significantly influenced the investors' desire to move to the stock market, shifting a significant share of capital from banks to the stock market, which has an additional benefit of being able to use tax benefits through an individual investment account. It is important to note that over 4.2 million Russian investors have opened brokerage accounts since 2020. This trend has led to increased investor interest in investing in the most profitable and least risky assets as popular and in-demand income sources. This is why the need to develop new tools for determining the profitability and riskiness of investing in assets, as well as for forecasting changes in the main indicators in the stock markets, has increased.

Thus, this topic is drawing people's attention, which leads to the need to consider various factors affecting a company's stock prices. This paper considers the possible influences of social media and analyses certain words and statements in them that can influence stock prices (and, consequently, their profitability). It goes without saying that the internet does not show influence on stock markets. It is necessary to consider the behavior of investors in the market; said behavior indicates their potential reactions to different situations, such as panic selling, holding, or buying more when a stock's price is crashing. However, the fall or rise of securities prices is not the only characteristic indicator for an investors' actions. They are significantly influenced by data streams, such as mass media, communication with other traders, and insider information leakage, etc. It is important to understand what kind of "news" prompts investors to do this or that in the stock market, which is the main focus of this research. Investor behavior and the existing methods of their analysis are presented in the literature review.

Thus, it can be assumed that a companies' prices change in different directions over a certain period, and these changes are caused by a number of factors (Ho and Huang 2021). The development of information channels and the growing number of non-professional investors make it necessary to study the impact of the investors' behavior expressed in quantitative characteristics on stock prices. In other words, a sudden increase in the number of people who want to invest in stocks but cannot do it independently and professionally due to lack of experience and sufficient knowledge in this area; the emergence of numerous websites and social media where both professional and non-professional investors can share their opinions; and news sources publishing both reliable and unreliable information, necessitates studying the impact of the investors' behavior on stock prices in order to evaluate whether a stock price change is justified or caused by falsified opinions and news. Thus, investors, having unique personalities and psychological characteristics, tend to react one way or another to stock market changes, and are an important element influencing stocks prices (Liapis et al. 2021). Taking recent events into account, investors are increasingly interested in alternative investment options, which require the skill of analyzing stock market trends and predicting the most likely changes. Additionally, the market is full of inexperienced inventors acting irrationally, which is why it is especially important to understand the psychology of investors and anticipate their actions in response to market events. Emotions and individual character traits can lead to significant losses and market changes, which brings in the need to develop tools that would be able to forecast the probability of stock quote changes (Sevumyan 2021). Besides, rational restrain artificially induced price movements by following subjective estimations uninfluenced by websites, but the discussion in social media provoking non-professional investors to act in one way or another promotes strong movements of stock quotes, which cannot be corrected by the behavior of rational investors (Xiong et al. 2020).

Despite this, most investors analyze market and company information before buying stocks. Market information is available on social media as well as in the news, blogs, and customer reviews of companies (Mndawe et al. 2022). Investors look at company profiles, historical trading data, news, analysts' opinions, financial reports, and other reports, to determine which company's stock value has increased or decreased (Bostan et al. 2020). Manual analysis of this information is prone to error, so machine learning techniques are used to find relationships between parameters (investor sentiment and stock prices) and predict the future values of dependent variables (stock prices). Javed Awan et al. (2021) uses various machine learning techniques to confirm that there is a significant and meaningful relationship between stock prices and data published in information sources and to justify the need to use machine learning techniques to analyze dependencies and predict stock prices. Jiao et al. (2016), as well as Li et al. (2020), and Zhang et al. (2018), address regression analysis and models where the independent variable is investor sentiment, and the dependent variable is stock price. Regression analysis of many studies has shown that a high determination coefficient provides a large percentage of the explanation of price changes under the influence of the information environment, represented by investor sentiment (messages, news). It is important to note that the mentioning of companies' names and the use of certain words (tokens) and expressions also carry a certain message, which is sure to be reflected in investor sentiment. In other words, the number of references to certain companies, securities, and stock market events, is also related to stock price movements that are in some way related to a company, or an event, etc. from the news. Thus Jiao et al. (2016) determined that a high discussion volume in social media predicts higher future volatility, and this relationship is statistically significant. It is important to note that a high volume of social media discussion causes higher volatility than a high volume of mass media discussion. However, it is not only the mentions of a company that affects their future stock prices, but also mentions of related companies, events, or shareholders; works using related words have more accurate results in determining the impact of investor sentiment on stock prices (Li et al. 2017; Mendoza Urdiales et al. 2021).

Li et al. (2020) delve into the effect of investors' comments on stock prices using long short-term memory (LSTM), the support vector machine method, and Bayes naive model, and conclude that daily investor sentiment contains predictive information for opening prices only, while hourly sentiment has the most accurate prediction for closing prices. In addition, there is also the idea that the more news is published, the more active the buying/selling of assets in the stock market is; however, it is also noted that buying is more active. Studies that divide investor sentiment into positive and negative have concluded that pessimistic investor sentiment affects quotes more than optimistic, and companies with optimistic investor sentiment have significantly higher stock returns in the current month than firms with pessimistic sentiment (Fang et al. 2021). In addition, there are also papers confirming that the relationship between publications on investor communication websites and average weekly stock returns mainly refers to periods when investor communication was inactive, and that higher stock returns do not follow investor sentiment (Xiong et al. 2020).

An important question is, who exactly has a significant and meaningful impact on investor behavior and, consequently, stock prices? For example, some investors believe that statements by opinion leaders have a strong influence on the masses, causing new trends in the stock market, which provides the following factors that determine the strength of the message: (1) the recipient's factors (the recipient's personality traits that characterize the degree of exposure to social media); and (2) properties of technology (to the technological differences of social media) (Vladimirovna and Quentin 2016). In continuation of the question of who, or what, has a greater influence on investor behavior and stock prices, there is an assumption validated using machine learning techniques that publications from users with more subscribers have a greater influence on same day returns, while publications from users with few subscribers have a greater influence on future returns (Sul et al. 2016). Authors of other works are of the opinion that it is not necessary to give people false information to induce them to perform a certain action or examine the authority of sources, the number of their subscribers, as the same can be achieved by simply finding accomplices who will consciously act in a way that objectively should not be acted upon. This idea is illustrated by an example when a Telegram channel posted a news story saying that it was now recommended to buy stock of Rospadskaya Coal Company, causing the stock price to grow and reach the highest value in the last 8 years, even though there were no justifiable reasons to buy this company's stock, other than the positive information given to the channel's subscribers. Such artificial interventions in the stock market most often occurred due to the saturation of the stock market in 2020–2021 with non-professional investors who cannot distinguish between "genuine" information and manipulation, due to their inexperience (Zvyagintseva and Ovchinnikova 2021).

An important note is that the existing methodologies for analyzing the impact of the information environment (used in the works described above), in particular investor sentiment on social media and on stock prices, are mostly similar; however, recently (in 2020 and 2021), published studies have improved the tools used due to the growing interest in this topic (Oliveira et al. 2021).

Many of the papers collected primary data in one way or another. Thus, investor websites, news resources, and social networks (Twitter) were analyzed. Comments, tweets, and news, are often parsed automatically using the Python 3 programming language and a number of libraries. Machine learning (ML), a field of computer science by which computer systems can make sense of data in the same way that humans do, is used to analyze data, most commonly for predicting stock prices. Simply put, ML is a type of artificial intelligence that extracts patterns from raw data using an algorithm or method. The most common programming languages used to create ML programs are R, Python, Scala, and Julia (Machine Learning 2021). The works mentioned above used such ML methods as: decision tree (classifying objects using questions about their attributes and their answers, moving along "branches" from each branch until the last question is answered); random forest (creating a "committee" of decision trees with different sets of attributes, forming

different classes, with the results of each tree, and selecting the most likely class); and clustering (grouping data items that have similar characteristics with statistical algorithms).

Figure 1 presents one possible model for collecting and analyzing social media data for market forecasting in most studies, evaluating the impact of investor sentiment on stock prices.

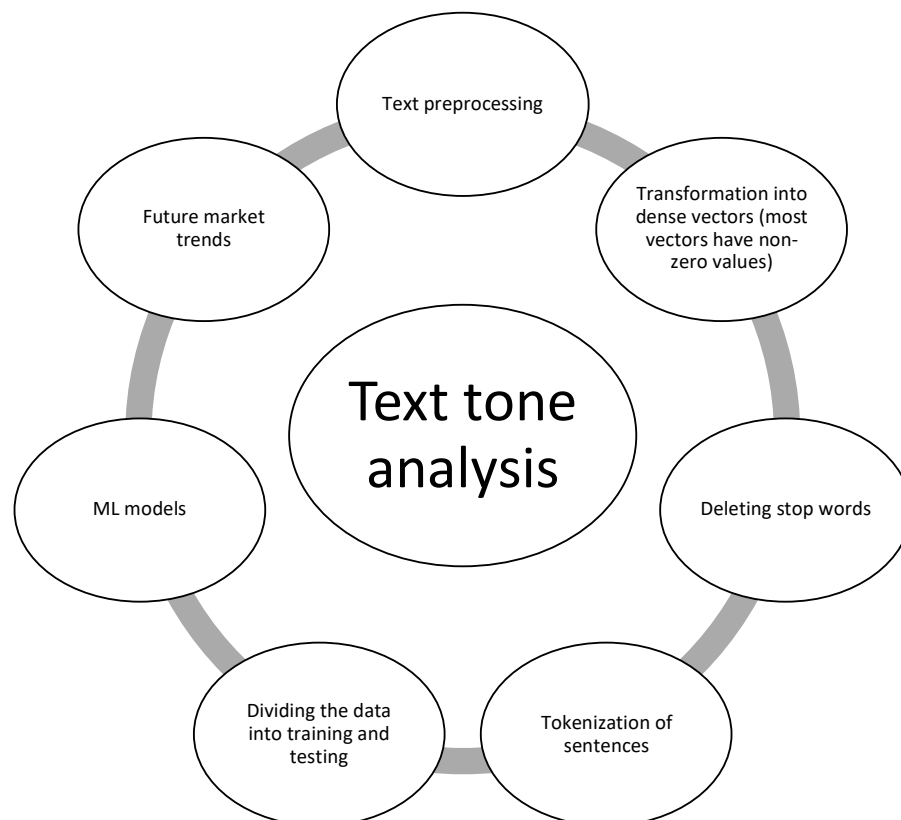


Figure 1. Conceptual model describing the impact of investor sentiment on stock prices.

Deep learning is a ML method that trains a neural network to predict outcomes from an input dataframe (a structure in which data is stored in a table). Advanced deep learning models such as long short-term memory networks (LSTM) are capable of capturing patterns in time series data, hence it is possible to apply them to predict future data trend ([Predicting Time Series with LSTM in Python 2021](#)). Recent research introduces a deep learning model called long short-term memory (LSTM) to make stock return and commodity price predictions ([Yurtsever 2021](#); [Alkhatib et al. 2022](#); [Sako et al. 2022](#)).

Almost all of the analyzed studies use a naive Bayesian algorithm (a classification algorithm based on Bayes theorem with the assumption of independence of attributes) to analyze investor sentiment ([Gamal et al. 2018](#)). Messages, news, and tweets are classified into three categories: negative (about a stock decline); positive (about a rise); and neutral (no change). Negative and positive tweets are applied to construct a sentiment index at some time interval, then the correlation of the index and stock price changes is calculated. In some works, as shown by Limongi [Limongi Concetto and Ravazzolo \(2019\)](#), this algorithm is applied to classify messages into three categories: buy; hold and sell or bull; neutral and bear.

Apart from directly highlighting expressions about the expectation of a stock quotes' decrease or increase in the message, it is also necessary to highlight their emotional component. Thus, the tone of the message is investigated by means of intellectual text analysis. However, some studies either do not consider the tone or are analyzed by special financial analysts (only the content component is considered, without the emotional component). To analyze the tone, the authors of some research articles used Opinion Finder (a system

that processes documents and automatically identifies subjective sentences, measuring positive and negative mood) and Google Profile of Mood States (an algorithm that measures mood and classifies it into six parameters: happiness, kindness, alertness, sureness, vitality, and calmness).

Thus, most works that consider the relationship between the content of investors' text messages and stock price movements contain similar tools for estimating these parameters and predicting future trends (Mendoza-Urdiales et al. 2022). It is also important to note that all the ways to assess the impact of investor sentiment on asset prices, in one way or another, come down to collecting data on price movements, message, publication, and tweet texts, then analyzing their emotional and content component. It is the parameters of prices and the tone or frequency of mentioning words and their correlation with prices that are the main components of all studies, including this paper. Tone is a textual category, which reflects the author's emotional attitude in achieving a specific communicative goal, and their psychological position in relation to the text presented, as well as to the addressee and the circumstances of communication (Alekseeva et al. 2011). Tone analysis is a field of computer linguistics, which deals with the study of opinions and emotions in textual documents. Its purpose is to find emotionally colored words and word combinations in a text and relate them to particular tone classes. The number of tone classes for a particular text is determined by the researcher applying tone analysis. Often, the task of binary tone classification is considered, i.e., there are only two classes: "positive" and "negative". As stated above, in the message we can analyze the author's emotion or the theme of the text itself. Thus, breaking down the text into individual words and working with them can also help in determining the relationship between investor sentiment, expressed in the use of certain words, and stock prices. For example, a direct mentioning of the word "decline" in a message in the investor chat may indicate a decline in stock prices, because some non-professional investors, trusting the information source and not having the skills to analyze the stock market events, may sell their stocks, which would actually contribute to the decline of stock prices (Zhang et al. 2022). However, it is necessary to perform pre-processing before analyzing the text. The following steps for this stage must be highlighted:

- Tokenization is the splitting of sentences into words;
- Removal of unnecessary punctuation, tags;
- Removing stop words (frequently occurring words that do not have definite semantics);
- Stemming—words are reduced to the root by removing the inflection and unnecessary characters, usually the suffix;
- Lemmatization is another approach to eliminate inflection by identifying parts of speech and using a detailed language database. Classical ML approaches ("Naive Bayes" or "Support Vector Machine") are suitable for further work (Pak 2012).

2. Methodology

After the above steps, the text will be suitable for conversion into a numeric format to continue the extraction of features (quantification of the resulting list of words). Further, it is possible to compile different indices, calculate the frequency of the presence of particular words, and their correlation with the movement of stock prices (Gayakwad et al. 2022).

Based on the parameters described above, the conceptual model of the theme can look as follows (Figure 2):

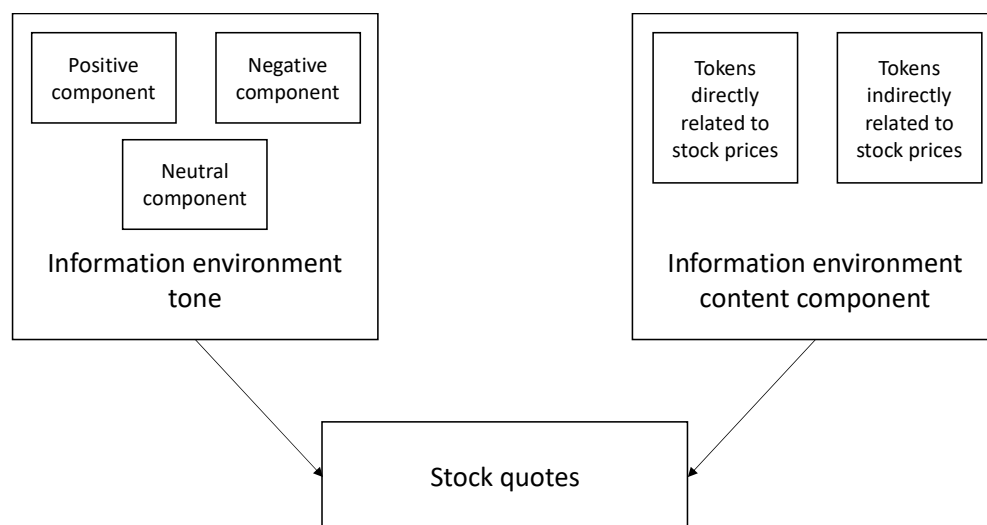


Figure 2. Conceptual model describing the impact of investor sentiment on stock prices.

Thus, the feasibility of examining different stocks influenced by the information environment has been analyzed. A literature review of existing studies on the topic of investor sentiment impact on stock prices has confirmed the interconnection and impact of professional and non-professional stock market participants' messages on stock prices. Existing techniques are based on machine learning methods that allow us to significantly simplify the processing of information. That is why the quantification of the information environment, namely the numerical evaluation of the emotional component of the environment participants and individual tokens (the content component) is considered as the basis for studying impact in the conceptual model. Thus, the conceptual model is based on the consideration of the influence of both positive, negative, and neutral components of the information message, as well as the mention of individual words and their frequency, on stock prices, which, in turn, ensure the investment attractiveness of the company/asset.

In order to assess the presence of the topics of events on the market and the frequency of references to certain securities, it is necessary to represent these content-thematic components by a token array. Consequently, in parallel with the tokenization of arrays of text information describing the information environment of financial markets, it is necessary to tokenize an array of text information describing actual events and the dynamics of changes on financial markets.

Within a consolidated news stream, a separate news stream of financial market events must be identified, from which news containing tokens describing the nature of those events, and the attitude of the news source/investor to that event, must be identified.

In one coefficient—financial market topics presence coefficient—it is proposed to identify such tokens as “stocks”, “bonds”, “futures”, and “currency”. The full range of tokens should be determined with the help of an associative algorithm, the use of which will provide the search of relationships (associations) with the main (presented above) tokens, which will help provide the most complete and clear picture formed in the news flow aggregate. Using a predefined list of tokens and the researcher's imperfect human knowledge will lead to distorted results, while using an associative series will fill the “gaps” in the predefined list of tokens that describe only a part of the information that the researcher is interested in. Thus, the tokens used in this coefficient describe: the presence of financial market topics in the information environment of the news flow that allows us to determine the “discussion volume” of the suggested topic at a given moment; and to trace the dynamics of the coefficient change by comparing its values with the key financial market indicators, in order to find out whether the correlation between the ratio movement and the main financial market parameters is present. Coefficient interpretation can be

represented as follows: at a certain point of time t the presence of financial market topic in the information environment of the news flow amounts to the received share (percentage).

To analyze the tone of the information environment it is necessary to consider the tone base coefficient, which characterizes the ratio of negative and positive tone. Its method of calculation suggests the following interpretation of the received value: the received value of the tone base coefficient reflects the ratio of negative and positive tokens at a certain moment of time (Koukaras et al. 2022). If the coefficient is greater than 1, we can say that negative judgments, opinions, and forecasts about the dynamics of events prevail in the information environment of the financial market, expressed by news units (opinions of experts and investors) at a certain point of time. If the coefficient is less than 1, the information flow reflects the predominance of tokens characterized by positive tone. The consideration of the tone coefficient in dynamics provides a researcher with an opportunity to make a conclusion about the direction of the emotional state of the participants and the authors of the information flow, which allows for making appropriate conclusions about the changes in the prices of financial instruments and, consequently, their volatility. The more the tone base coefficient fluctuates for a period, the more investors can be confident about the unstable situation in the securities market, and, consequently, the number of speculators increases. Thus, the coefficient approaching a zero value over time is a kind of signal for investors about a stable positive information environment, which allows for reducing the risk of losses.

The next indicator evaluating the tone of the information environment is the negative tone class dispersion. In a general case, the dispersion characterizes the mean value of the square of the deviation of a random variable from the mean. Adapting the concept of dispersion to the ongoing study, the author suggests the following understanding of this indicator: the mean value of the squared deviation of the information environment's negative tone from the mean value for the period under consideration (Kanavos et al. 2022). The larger the value of the negative tone class dispersion, the larger the possible change of "negativity" of the information environment both towards the increase of negative tone, and towards the decrease. The high values of this indicator widen the zone of uncertainty in making investment decisions, and, accordingly, with the reduction of the negative class dispersion values the investor can judge about the reduction of uncertainty in future values of the main indicators of financial instruments, which reduces the possible risk of financial losses. The calculation methodology implies the following interpretation of the obtained results: the obtained value reflects the average square of the deviation of the negative tonal component of the information environment from the average value for the period under consideration (Chun and Jang 2022).

Thus, three coefficients will be considered, two of which characterize the tonal content of the information environment and one coefficient, fundamentally describing the entire content component: the tone base coefficient; the dispersion of negative tone class; and the coefficient of financial market topics presence. The set of such coefficients developed by the author, assuming meaningful results and their interpretation, is presented in Table 1.

The described coefficients require extracting news flow from the information environment of the financial market and tokenizing and quantifying news arrays of natural information that characterize the information space of a potential investor. Calculation of coefficients specified in Table 1, information environment state analysis, and its tonal and content components, allows for designing a universal automated algorithm of their stage calculation, using programming language Python 3.

In the author's opinion, to identify the basic conclusions in the financial market information environment quantification, it is necessary to study large companies represented in the global market, mentions of which are often found in the news, i.e., they are most clearly represented in the information environment. Firstly, these are the companies whose shares are "known" to most people and are particularly attractive to unqualified novice investors. Secondly, these are companies that regularly make changes to their operations and products and announce it on the Internet. Third, these companies are online marketplaces. Thus,

to analyze the impact of the proposed information environment indicators on the main parameters of the financial market, the largest international companies whose presence in the information environment of the global financial market is estimated to be high, were selected as an example—Tesla Inc., Apple Inc., and Amazon Inc. These companies were chosen because of their significant representation in the information environment, which will ensure that there are not many null values of the information environment indicators chosen by the author of the study, due to their significant presence in news streams. In addition, it is important to consider how the information environment affects the dynamically developing company Tesla Inc., opinions on which are polarized; the stable company Apple Inc., which is an IT giant with a huge capitalization and arouses significant interest for investment among “newcomers” to the financial market; and the largest marketplace and global e-commerce giant—Amazon. com Inc., which has particular appeal for investors, had phenomenal stock price growth in 2020, and the future prospects of which are being discussed by hundreds of experts in the financial market.

Table 1. Set of information environment quantifiers.

Factor Name	Description	Formula
Tone base coefficient	Characterizes the ratio of negative tokens to positive tokens. The calculation methodology assumes the following interpretation of the obtained value: the obtained value of the tone base coefficient reflects the share of negative tokens in the positive ones at a certain moment of time. The consideration of the coefficient’s dynamics provides the researcher with an opportunity to make a conclusion about the direction of the emotional state of participants and the authors of the information flow, which allows making appropriate conclusions about the changes in the prices of financial instruments and, consequently, their volatility.	$K_{TBi} = \frac{T_{neg_i}}{T_{pos_i}}$ T_{pos} —optimistic market evaluation. T_{neg} —pessimistic market evaluation.
Negative tone class dispersion	Shows the mean value of the squared deviation of the information environment’s negative tone from the mean value for the period under consideration. The greater the value of negative tone class dispersion, the greater the possible change of “negativity” of information environment both in the direction of negative tone increase, and in the direction of decrease.	$K_D = \frac{\sum_{i=1}^n (T_{neg_i} - \overline{T_{neg}})^2}{n}$ K_D —negative tone class dispersion. n —number of tokens analyzed T_{neg_i} —pessimistic market evaluation. $\overline{T_{neg}}$ —mean pessimistic market evaluation.
Financial market topics presence coefficient	By calculating the share of “stocks”, “bonds”, “futures”, “currency” tokens in the total aggregate of tokens representing the news flow of the information environment, it describes the presence of financial market topics in the information environment of the news flow, which allows to determine the “discussion volume” of the proposed topic at a particular point in time, trace the dynamics of the ratio change, comparing its values with the key financial market indicators in order to identify the presence of correlation between the ratio movement and key financial market parameters. Interpretation of the coefficient can be represented as follows: at a certain point of time t the presence of financial market topic in the information environment of the news flow is the received ratio (percentage).	$K_{FMi} = \frac{m(T_{FMi} \cap T_i)}{m(T_i)}$ K_{FM} —share of tokens related to financial markets topics in the news stream. $m(T_{FM_i})$ —array of tokens related to financial markets topics in the news stream. $m(T)$ —array of all tokens in the news stream.

Stock prices (in dollars) of these companies were collected from the American stock exchange, Nasdaq, and their volatility was calculated for the period of 1 April 2020 to 31 August 2020 according to the Formula (1):

$$V = \frac{P_{max} - P_{min}}{P_{min}} * 100\% \tag{1}$$

where P_{max} is the maximum share price in 1 day, USD, and P_{min} is the minimum share price for 1 day, USD.

The reason for choosing the period of 1 April 2020 to 31 August 2020 is the special influence of information flow in the financial environment during the pandemic period, when the stock market was flooded with private, unqualified investors, the information environment was characterized by the creation of a huge number of investor chats, and people decided to find new ways to earn money.

The analysis of influence of information environment on financial market parameters is carried out by means of regression analysis with application of Excel program possibilities, namely the package of analysis tools—data analysis. Table 2 presents a summary table of indicators.

Table 2. Summary table of indicators.

Indicator Name	Unit Designation	Unit of Measure	Type	Source
Tonal base coefficient	K_{TB}	%/ratio	exogenous	Google News
Negative tone class dispersion	K_D	%/ratio	exogenous	Google News
Financial market topics presence coefficient	K_{FM}	%/ratio	exogenous	Google News
Share price	Y_P	\$	endogenous	Nasdaq
Share price volatility	Y_V	%/ratio	endogenous	Nasdaq

For regression analysis with the help of an automated information environment quantification algorithm, we will calculate the values of quantifiers from Table 1, clean the obtained values of the coefficients characterizing the information environment from outliers, and delete the data on the days with zero values of the coefficients. The reliability level chosen for the model is 90%. It is important to note that the information environment affects the company stock prices with a lag of 1 day. Optimization of regression equations allows us to draw the following conclusions:

1. In case of Tesla Inc. when excluding variables with p-level greater than 10%, in the stock price dependence on information environment indicators F-criterion significance was 42.4%, which is more than 10%, which means that dependence of financial market parameters on information environment indicators is insignificant. In addition, the determination coefficient in the same dependence is only 6.85% (less than 40%), which does not meet the above constraints, and the approximation error is 35.55%, which is much higher than the permissible value (15%). When optimizing the volatility equation of Tesla Inc. stock price volatility, the determination coefficient is 23.48% and the approximation error is 54.30%, which is significantly higher than the proposed constraint.
2. In case of Apple Inc. when excluding variables with p-level greater than 10%, in the stock price dependence on information environment indicators the determination coefficient was 16.55% (less than 40%), which does not satisfy the above restrictions, and the approximation error was 15.63%, which is significantly greater than the permissible value (15%). When optimizing the volatility equation of Apple Inc. stock price volatility, the F-criterion significance was 24.19%, which is more than the permissible value (10%) the coefficient of determination was 20.44% (less than 40%) and the approximation error was 35.06%, which is significantly higher than the proposed constraint.
3. In case of Amazon Inc., when optimizing the volatility equation of the share price volatility of Amazon Inc. the determination coefficient was 22.4% and the approximation error was 31.19%, which is significantly higher than the proposed constraint. The regression analysis of the stock price of Amazon Inc. proved the feasibility of

researching the effect of information environment indicators on the stock prices of Amazon Inc.

3. Results

Thus, it was decided to exclude from further analysis the regression equations of information environment indicators' influence on stock prices and their volatility for Tesla Inc. and Apple Inc. and the influence of information environment indicators on volatility of Amazon Inc. stock. The following study evaluates the accuracy and plot analysis of the theoretical and actual values, residuals of the regression equation of the impact of information environment quantifiers on the stock price of Amazon Inc.

The results of regression analysis and optimization of regression equations by p-level, namely, determination coefficient, F significance and error of approximation are presented in Table 3.

Table 3. Results of regression analysis after optimization of regression equations by p-level, %.

The Company	Dependent Variable	F Significance	R ²	A
Tesla Inc.	Share price	42.40	6.85	35.55
	Share price volatility	0.58	23.48	54.30
Apple Inc.	Share price	4.13	16.55	15.63
	Share price volatility	24.19	20.44	35.06
Amazon Inc.	Share price	0.54	49.24	8.56
	Share price volatility	6.83	22.4	31.19

Table 4 presents an equation describing the influence of information environment quantifiers on financial market indicators.

Table 4. Equation describing influence of information environment quantifiers on financial market indicators.

The Independent Variables in the Equation	Coefficient
Constant	1879.75
K_{TB}^2	2012.29
K_{TB}^3	−2382.43
K_{TB}^4	733.46
K_{TB}^5	−0.40
K_D^2	443.76
K_D^3	−116.54
K_D^3	1.32
K_{FM}^3	−2,286,999,084.24
K_{FM}^4	268,452,154,575.78
K_{FM}^5	−10,660,324,751,915.70
K_{FM}^6	141,229,550,932,683.00
Resulting Equation	
$Y_P = -0.40 \cdot K_{TB}^5 + 733.46 \cdot K_{TB}^4 - 2382.43 \cdot K_{TB}^3 + 2012.29 \cdot K_{TB}^2 + 1.32 \cdot K_D^5 - 116.54 \cdot K_D^3 + 443.76 \cdot K_D^2 + 141,229,550,932,683 \cdot K_D - 141,229,550,932,683 \cdot K_{FM}^6 - 10,660,324,751,916 \cdot K_{FM}^5 + 268,452,154,576 \cdot K_{FM}^4 - 2,286,999,084 \cdot K_{FM}^3 + 1879.75$	

Figure 3 shows a chart comparing actual and theoretical stock prices of Amazon Inc., while Figure 4 shows a residual plot.

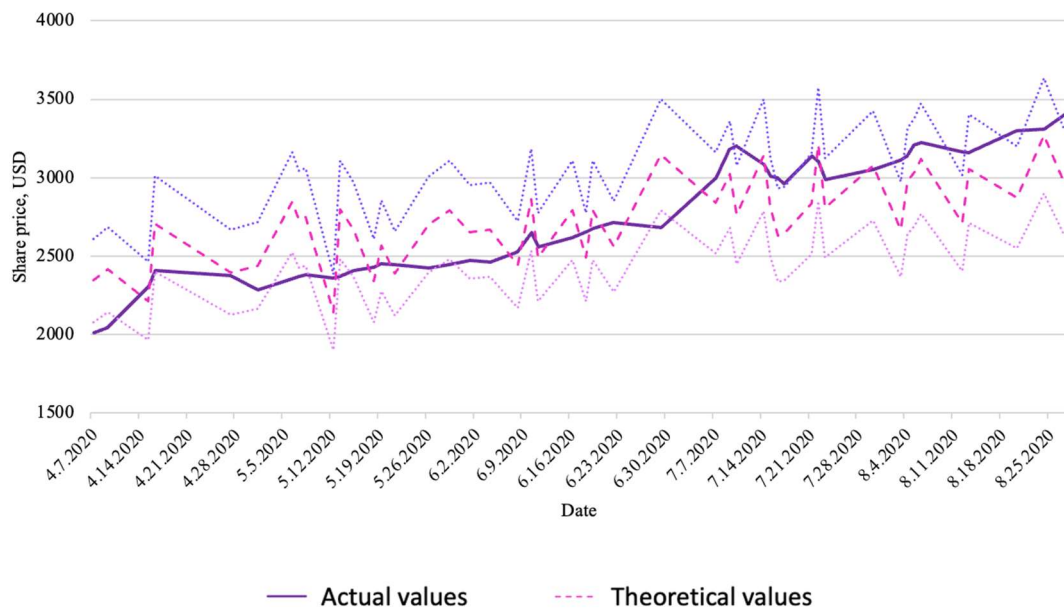


Figure 3. Graph comparing actual and theoretical values of Amazon Inc. stock prices.

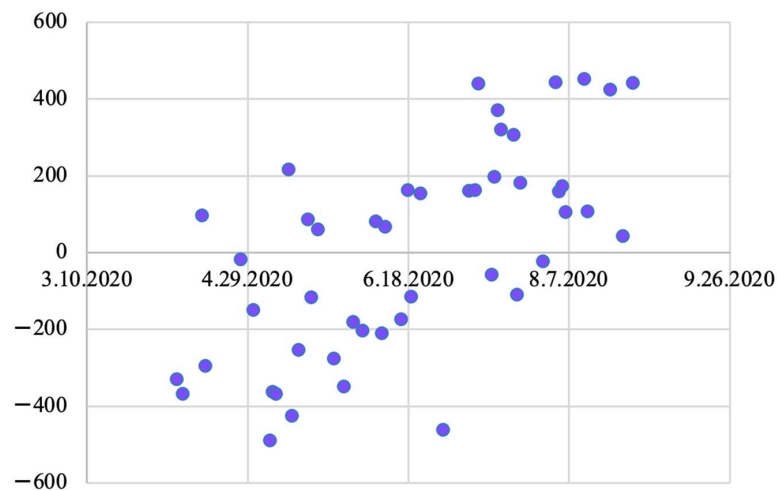


Figure 4. Residual plot.

When first viewing the chart, our attention undoubtedly stops at the fact that the actual share price value of Amazon Inc. is almost always below the theoretical value until 8 June 2020. Consequently, the past systematically points to a potential increase in the dependent variable. Obviously, there are systemic factors that exert a near-constant understatement of the actual value over the theoretical value (prior to 8 June 2020). In other words, the proposed equation less accurately describes the dynamics of prices in the period from 7 April 2020 to 8 June 2020, which is most likely due to the presence of other factors influencing the price movement of the company.

From 7 April 2020 to 16 April 2020 there was a significant increase in the company’s share price, which was provided by the increased demand for Amazon services, which represent several segments: Amazon Inc.’s own online sales; offline product retailing; services that Amazon provides to companies selling goods through its platform (logistics and marketplaces); company products operating on the subscription model; and the company’s cloud business, etc. In particular, the increase in demand is due to the lack of free movement during the period of coronavirus infection and the popularity of online services, which is the main focus of Amazon. However, the growth of its stocks is provided

by the fact that in early April 2020 it announced that it was hiring 75 thousand new employees (the day before it had already hired 100 thousand people to its distribution centers) (Kirichenko 2020). This is a positive signal for investors because the extensive development of the company manifested in the hiring of new employees, indicating the expansion of the company, and therefore the success of its operations, and, accordingly, the volume of Amazon stock trading increased, subsequently raising the stock prices.

From 16 April 2020 to 1 May 2020 there is a slight decrease in the company's stock prices. If at the beginning of the above-mentioned period some value of quotations is provided by the equalization of private investors' demand for stocks of Amazon, Meta, Facebook, etc., since 27 April 2020, a significant decline is provided by the fact that Amazon had reported last quarter sales of \$75.5 billion, up 26 percent from a year earlier, beating analysts' expectations. Profits fell 29 percent to \$2.5 billion, worse than expected, because it cost more to meet increased consumer demand. Amazon CEO, Jeff Bezos, made it clear that profits could continue to fall in the near future. Such news may have demotivated investors to buy Amazon stock (Weise 2020).

Then, from 5 June 2020 to 29 June 2020, there was a relatively steady rise in the company's share price with a bounce on 11 June 2020, which could be explained by the announcement of the news that Amazon would hold the first Singapore Seller Summit to help local businesses take advantage of growth opportunities online. The free online event was designed to bring together the local seller community and enable them to develop strategies and skills to expand their business and attract more customers locally and globally. This news undoubtedly ensured an influx of Singapore investors and an increase in Amazon's share price.

From 29 June 2020 till 27 August 2020 there were some growth spurts, and from the end of July 2020 growth was already stable. The growth spurt of 7 July 2020 to 17 July 2020 was due to the continued growth of online sales and their popularity, which increased the market capitalization of the company and led to an influx of investors. In addition, in July 2020, Amazon's upcoming entry into the Russian market became known (creation of a joint cloud platform (MCS) with Mail.ru) (Smirnov 2020). This prompted investors to buy the company's shares, as the company was expected to grow after entering the Russian market. Moreover, in June–July 2020, Netflix raised prices for its subscribers, while failing to increase their number, which struck a blow to Netflix's performance and provided an influx of new users of the Amazon Prime service. However, according to experts, it was also a negative signal for Amazon, as Netflix was among the five largest and most successful U.S. technology companies along with Facebook, Amazon, Apple, and Google. A drop in Netflix shares could signal to market participants to sell stocks of the other four companies. Perhaps that is why there was a decline after a growth spurt (Tsegoev 2019). Further growth was provided by the stable position of the company in the market and the popularity of all sectors represented in Amazon.

Let us consider the graph in Figure 3 in more detail. A sufficiently large number of structural breaks and outliers are observed. Thus, structural outliers (outliers of actual values beyond the lower and upper limits of the permissible range) are observed on 7 April 2020 to 9 April 2020, 6 May 2020 to 8 May 2020, 13 May 2020, 29 May 2020, 29 June 2020, 10 July 2020, 16 July 2020 to 17 July 2020, 3 August 2020, 12 August 2020, 20 August 2020, and 27 August 2020. Structural outliers at the beginning of April 2020 are due to the stock market crash in March 2020 on the back of the coronavirus pandemic that began, and political and economic relations that the model for the impact of information environment indicators on financial market performance does not take into account. Further outliers are also mainly caused by factors not considered by the author of this study. Structural outliers account for 32% of all events considered. Thus, during the pandemic, many companies adapted to the new realities of the world, and there was government intervention in the economy (subsidies to businesses). Developments in the financial market, and the U.S. market in particular, also influenced the behavior of unqualified investors (e.g., fear of contagion, falling markets at the beginning of the pandemic prompted unwarranted buying

and selling of company shares), which also created structural outliers. It is important to note that in the first half of the period under consideration, actual values were mainly below the values of the lower bound of the allowable interval, while in the second half Amazon share prices predominantly exceeded the upper bound. This is evidenced by the residual chart in Figure 4. This may indicate that during the first half of the period Amazon stock was underpriced, while during the second half it was overpriced. Importantly, these events are not so vividly represented in the information environment. Thus, in the first half, undervalues (negative residue) were caused by the financial market crash and by investors' caution in further purchases of securities, while in the second half the positive residue was caused by the popularity of online sales, the development of cloud technologies, and in Amazon's support to many small and medium businesses to enter online sector, which contributed to investing in the company due to increased confidence and expectations of growth of quotes, as well as the publication of positive financial statements on the expansion of operations, capitalization, and profit growth.

Besides, we should pay attention to the structural breaks that were observed in 45% of the considered events. However, it is incorrect to compare changes of quote values and their prefix (+/−) to outliers, as some information is lost in this analysis and the structural breaks with outliers may give wrong results due to false prefix changes.

The model can be called optimal as it takes into account all the constraints and in 68% of cases lies in the acceptable interval. The presence of heteroscedasticity in the residuals suggests the presence of other factors affecting the formation of Amazon stock prices, which does not contradict the author's hypothesis. Obviously, the presence of systemic factors distorts the indicators of the information environment on Amazon stock prices in the graph, but compliance with the other criteria gives us an opportunity to trust this model with 90% probability.

4. Discussion

As a result of the analysis of the influence of indicators characterizing the information environment of the financial market on the main parameters of the financial market, and establishing the basic criteria that ensure sufficient validity of the model and validity of financial market parameters to a sufficient extent, it was determined that the obtained regression equation is suitable by all criteria for explaining the share price movements only in the case of Amazon. In the author's opinion, this result is caused by a wide representation of Amazon in the online sphere, as the main services offered by this company are based on online sales. Amazon is not only a giant online marketplace, but also a global provider of cloud computing and artificial intelligence services. The wide range of services offered online ensures that the company's discussion is ubiquitous. According to a survey of Amazon customers by Cowen Bank, the coronavirus "has caused a significant increase in the share of online shopping due to the closure of physical stores and the need for social distancing". This fact increased the volume of discussion of the company in question, as various companies began to discuss the successes and mistakes of the company; also at this time, the company itself actively published news about its activities.

At the same time, it is important to note that despite the fact that Apple was also performing well in the market at that time, its discussion volume reached a significant level of impact on its stock quotes. Although the company did publish important news about the migration of Apple computers to Apple's own Apple Silicon processors, and the announcement of the release of iOS 14, this did not significantly affect the information background around it. Apple is rather a steadily developing company that has gained enormous popularity, which is unlikely to ever decrease its turnover on the market due to the innovations introduced in its products and the high quality of the goods produced. The discussion in the information environment is rather cyclical and is not subjected to different opinions of unqualified investors, but it is developing due to the growing popularity of Apple Music, iCloud, and Apple Pay services, due to the increase in the amount of time spent at home and the need to develop contactless payment.

As for Tesla, its rapid growth and quote changes are mainly driven by changes in the automotive sector's commodity base, the actions of Elon Musk (selling/buying the company's stock), and the "Tesla financial complex"—"a phenomenon that many investors, whether those are passive index funds, traditional mutual funds, hedge funds or simple retail investors, cannot fight, given the unique strength it now has in the stock market" (Bekmagambetova 2020). Despite it having a fairly high discussion volume in the information environment due to its rapid growth and significant presence in the global stock market, the indicators characterizing the information environment of the company cannot provide an explanation for the fluctuations of the stock forces, because they occur under the influence of other factors mentioned above. In other words, fluctuations of the financial market information environment around Tesla do not explain fluctuations of the stock quotes, as they are mainly provided by the "Tesla financial complex".

5. Conclusions

This article has shown how investment decisions are taken on the basis of the information environment passing through the consciousness of the investor. In particular, it was proved that tools allowing us to assess the properties of the information environment in the context of the investor's consciousness make it possible to predict the vector of decisions he makes, as well as for larger market players, to influence market changes through information environment management.

Based on the obtained results, presented in the form of parameters of the regression model, a specialist can determine the most effective decisions regarding the acquisition of assets in the market, guided by the results of the quantification of the sentiment of the information environment. In addition, subjects that are to some extent capable of influencing the information environment (news aggregators, information portals (e.g., online newspapers), as well as authorities regulating their activities) can also be guided by the resulting model in making their management decisions.

In general, it can be concluded which companies can be subject to analysis using the algorithm proposed by the authors for calculating the system of quantifiers of the information environment of the financial market, and further regression analysis with a sufficient level of reliability and determinacy. In the authors' opinion, these are the companies that, in addition to sufficient discussion in the news, have online sales of daily and non-daily demand, and have a wide representation in the online sphere. These may be the following segments: e-commerce, cloud computing, artificial intelligence, online entertainment services, global online stores, and marketplaces.

The limitation of this study is that its results are based on the quantification of the information environment, limited to a single source (Google News). Also, the limitation of the study is that its results were obtained on the assets of companies widely presented in the media. This fact determines the presence of a large amount of related news in the information environment and a significant number of non-professional investors who are guided to a greater extent, or exclusively, by the information environment when making their own investment decisions.

As a prospect of this study, the authors can name the expansion of the list of analyzed assets (companies), which will allow us to classify the results by industries and types of activity, or other necessary classification criteria. This study can also be developed in the direction of including a content component of the information environment analysis in addition to the tonal one, as well as using more variable statistical tools (for example, neural networks), which can both improve the quality of the current model and expand its scope.

Author Contributions: Conceptualization, D.G.R. and E.A.K.; methodology, E.A.K.; software, O.A.K.; validation, P.A.P.; formal analysis, P.A.P.; investigation, P.A.P.; resources, E.A.K.; data curation, O.A.K.; writing—original draft preparation, E.A.K.; writing—review and editing, D.G.R.; visualization, P.A.P.; supervision, E.A.K.; project administration, O.A.K.; funding acquisition, D.G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program ‘Priority 2030’ [(Agreement 075-15-2021-1333 dated 30 September 2021)].

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All analytical data was obtained from open sources, revealing the value of assets in the securities market. The quantification of natural digital information was carried out on the basis of Google news resources.

Acknowledgments: We are grateful to Ministry of Science and Higher Education of the Russian Federation for given us the opportunity to participate in the academic leadership program ‘Priority 2030’.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Alekseeva, Ludmila, Annushkin Vladimir, and Bazhenova Elizabeth. 2011. *Stylistic Encyclopedic Dictionary of Russian Language*, 2nd ed. Moscow: LLC FLINTA.
- Alkhatib, Khalid, Huthaifa Khazaleh, Hamzah Ali Alkhazaleh, Anas Ratib Alsoud, and Laith Abualigah. 2022. A New Stock Price Forecasting Method Using Active Deep Learning Approach. *Journal of Open Innovation: Technology, Market, and Complexity* 8: 96. [CrossRef]
- Bekmagambetova, Dinara. 2020. How Do Tesla, Apple and Amazon Set New Records during the Pandemic. *Kursiv*. Available online: <https://kz.kursiv.media/2020-07-17/kak-tesla-apple-i-amazon-byut-rekordy-vo-vremya-pandemii/> (accessed on 15 March 2022).
- Bostan, Ionel, Ionela-Corina Chersan, Magdalena Danileț, Mihaela Ifrim, and Viorica Chirilă. 2020. Investigations Regarding the Linguistic Register Used by Managers to Convey to Stakeholders a Positive View of Their Company, in the Context of the Business Sustainability Desideratum. *Sustainability* 12: 6867. [CrossRef]
- Chun, Se-Hak, and Jae-Won Jang. 2022. A New Trend Pattern-Matching Method of Interactive Case-Based Reasoning for Stock Price Predictions. *Sustainability* 14: 1366. [CrossRef]
- Fang, Hao, Chien-Ping Chung, Yang-Cheng Lu, Yen-Hsien Lee, and Wen-Hao Wang. 2021. The impacts of investors’ sentiments on stock returns using fintech approaches. *International Review of Financial Analysis* 77: 101858. [CrossRef]
- Gamal, Donia, Marco Alfonse, El-Sayed M. El-Horbaty, and Abdel-Badeeh M. Salem. 2018. Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains. *Machine Learning and Knowledge Extraction* 1: 224–34. [CrossRef]
- Gayakwad, Milind, Suhas Patil, Amol Kadam, Shashank Joshi, Ketan Kotecha, Rahul Joshi, Sharnil Pandya, Sudhanshu Gonge, Suresh Rathod, Kalyani Kadam, and et al. 2022. Credibility Analysis of User-Designed Content Using Machine Learning Techniques. *Applied System Innovation* 5: 43. [CrossRef]
- Ho, Trang-Thi, and Yennun Huang. 2021. Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation. *Sensors* 21: 7957. [CrossRef]
- Javed Awan, Mazhar, Mohd Shafry Mohd Rahim, Haitham Nobanee, Ashna Munawar, Awais Yasin, and Azlan Mohd Zain. 2021. Social Media and Stock Market Prediction: A Big Data Approach. *Computers, Materials & Continua* 67: 2569–83. [CrossRef]
- Jiao, Peiran, Andre Veiga, and Ansgar Walther. 2016. Social Media, News Media and the Stock Market. *SSRN Electronic Journal*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755933 (accessed on 15 March 2022). [CrossRef]
- Kanavos, Andreas, Yorghos Voutos, Foteini Grivokostopoulou, and Phivos Mylonas. 2022. Evaluating Methods for Efficient Community Detection in Social Networks. *Information* 13: 209. [CrossRef]
- Kirichenko, Irina. 2020. Amazon: Stocks Boosted by Delivery, Cloud and Streaming TV. *Expert.Ru*. Available online: <https://expert.ru/2020/04/14/amazon-aktsii-povyishayut-dostavka-oblako-i-strimingovoe-tv> (accessed on 15 March 2022).
- Koukaras, Paraskevas, Christina Nousi, and Christos Tjortjis. 2022. Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning. *Telecom* 3: 358–78. [CrossRef]
- Li, Bing, Keith C. C. Chan, Carol Ou, and Sun Ruifeng. 2017. Discovering public sentiment in social media for predicting stock movement of publicly listed companies. *Information Systems* 69: 81–92. [CrossRef]
- Li, Yelin, Hui Bu, Jiahong Li, and Junjie Wu. 2020. The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting* 36: 1541–62. [CrossRef]
- Liapis, Charalampos M., Aikaterini Karanikola, and Sotiris Kotsiantis. 2021. A Multi-Method Survey on the Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting. *Entropy* 23: 1603. [CrossRef] [PubMed]
- Limongi Concetto, Chiara, and Francesco Ravazzolo. 2019. Optimism in Financial Markets: Stock Market Returns and Investor Sentiments. *Journal of Risk and Financial Management* 12: 85. [CrossRef]
- Machine Learning. 2021. Big Data School. Available online: <https://www.bigdataschool.ru/wiki/machine-learning> (accessed on 24 October 2021).
- Mendoza Urdiales, Román Alejandro, Andrés García-Medina, and José Antonio Nuñez Mora. 2021. Measuring information flux between social media and stock prices with Transfer Entropy. *PLoS ONE* 16: e0257686. [CrossRef] [PubMed]

- Mendoza-Urdiales, Román Alejandro, José Antonio Núñez-Mora, Roberto J. Santillán-Salgado, and Humberto Valencia-Herrera. 2022. Twitter Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods. *Entropy* 24: 874. [CrossRef]
- Mndawe, Sibusiso T., Babu Sena Paul, and Wesley Doorsamy. 2022. Development of a Stock Price Prediction Framework for Intelligent Media and Technical Analysis. *Applied Sciences* 12: 719. [CrossRef]
- Oliveira, Michele Gendelsky de, Graça Azevedo, and Jonas Oliveira. 2021. The Relationship between the Company's Value and the Tone of the Risk-Related Narratives: The Case of Portugal. *Economies* 9: 70. [CrossRef]
- Pak, Alexander. 2012. Teaching Computer Senses (Sentiment Analysis). Habr.Com. Available online: <https://habr.com/ru/post/149605> (accessed on 30 October 2021).
- Predicting Time Series with LSTM in Python. 2021. Pythobyte. Available online: <https://pythobyte.com/time-series-prediction-using-lstm-with-pytorch-in-python-521ce3ed/> (accessed on 24 October 2021).
- Sako, Kady, Berthine Nyunga Mpinda, and Paulo Canas Rodrigues. 2022. Neural Networks for Financial Time Series Forecasting. *Entropy* 24: 657. [CrossRef]
- Sevumyan, Elina. 2021. Forecasting of securities dynamics under conditions of investors' increased interest in exchange trade. *Финансовые Рынки и Банки (Financial Markets and Banks)* 6: 68–70.
- Smirnov, Anton. 2020. Is Amazon's Expensive Stock Worth Buying? Блог Ленивого Инвестора (Lazy Investor's Blog). Available online: <https://smfanton.ru/fondovaya-birzha/akcii-amazon.html> (accessed on 15 March 2022).
- Sul, Hong Kee, Alan R. Dennis, and Lingyao Yuan. 2016. Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns. *Decision Sciences* 48: 454–88. [CrossRef]
- Tsegoev, Vladimir. 2019. Stock Market Signal: How Netflix's Record Stock Collapse Threatens Major U.S. Technology Companies. RT. Available online: <https://russian.rt.com/business/article/651145-netfliks-obval-birzha> (accessed on 15 March 2022).
- Vladimirovna, Dubko Anna, and Boissier Quentin. 2016. Sustainable Influence of Social Media on Volatility of Stock Prices. State of Art. Cyberleninka. Available online: <https://cyberleninka.ru/article/n/sustainable-influence-of-social-media-on-volatility-of-stock-prices-state-of-art/pdf> (accessed on 15 March 2022).
- Weise, Karen. 2020. Amazon Earnings: Sales Are Up, but Company Warns of Higher Costs Ahead. The New York Times. Available online: <https://www.nytimes.com/2020/04/30/technology/amazon-stock-earnings-report.html> (accessed on 15 March 2022).
- Xiong, Xiong, Yongqiang Meng, Nathan Lael Joseph, and Dehua Shen. 2020. Stock mispricing, hard-to-value stocks and the influence of internet stock message boards. *International Review of Financial Analysis* 72: 101576. [CrossRef]
- Yurtsever, Mustafa. 2021. Gold Price Forecasting Using LSTM, Bi-LSTM and GRU. *European Journal of Science and Technology* 31: 341–47. [CrossRef]
- Zhang, Xi, Jiawei Shi, Di Wang, and Binxing Fang. 2018. Exploiting investors social network for stock prediction in China's market. *Journal of Computational Science* 28: 294–303. [CrossRef]
- Zhang, Zhishuo, Manting Luo, Zhaoting Hu, and Huayong Niu. 2022. Textual Emotional Tone and Financial Crisis Identification in Chinese Companies: A Multi-Source Data Analysis Based on Machine Learning. *Applied Sciences* 12: 6662. [CrossRef]
- Zvyagintseva, Natalia, and Kseniya Ovchinnikova. 2021. Securities Market Conditions: The Phenomenon of the Influence of Private Investors by Means of Social Networks. *Baikal Research Journal* 12. [CrossRef]