

Polak, Kamil

Article

The impact of investor sentiment on direction of stock price changes: Evidence from the Polish stock market

Journal of Banking and Financial Economics (JBFE)

Provided in Cooperation with:

Faculty of Management, University of Warsaw

Suggested Citation: Polak, Kamil (2021) : The impact of investor sentiment on direction of stock price changes: Evidence from the Polish stock market, Journal of Banking and Financial Economics (JBFE), ISSN 2353-6845, University of Warsaw, Faculty of Management, Warsaw, Iss. 16, pp. 72-90, <https://doi.org/10.7172/2353-6845.jbfe.2021.2.4>

This Version is available at:

<https://hdl.handle.net/10419/313450>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

The Impact of Investor Sentiment on Direction of Stock Price Changes: Evidence from the Polish Stock Market

Kamil Polak

Kozminski University, Warsaw, Poland

kamil.polak@kozminski.edu.pl

<https://orcid.org/0000-0002-9897-8431>

Received: 13 September 2021 / Revised: 8 December 2021 / Accepted: 18 December 2021 / Published online: 30 December 2021

ABSTRACT

The purpose of this research is to examine the impact of sentiment derived from news headlines on the direction of stock price changes. The study examines stocks listed on the WIG-banking sub-sector index on the Warsaw Stock Exchange. Two types of data were used: textual and market data. The research period covers the years 2015–2018. Through the research, 7,074 observations were investigated, of which 3,390 with positive sentiment, 2,665 neutral, and 1,019 negative. In order to examine the predictive power of sentiment, six machine learning models were used: Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, KNN Classifier, SVC and Gaussian Naive Bayes Classifier. Empirical results show that the sentiment of news headlines has no significant explanatory power for the direction of stock price changes in one-day time frame.

JEL Classification: G14; G17; G41

Keywords: sentiment analysis, natural language processing, machine learning, financial forecasting, behavioral finance.

1. INTRODUCTION

The dynamic technological development has significantly increased the information role of the Internet in recent times. The arrival of the era known as Web 2.0 gave rise to a completely new way of communication, namely social media. The exchange of information via online channels takes place almost immediately, which gives a significant information advantage over the traditional methods of data acquisition, such as board reports or articles. In hindsight, over the last few years social media platforms have become not only means of communication expressing their own opinions, but also publicly sharing emotions.

The development of technology has significantly contributed to the evolution of research tools and methods enabling the exploration of new areas, and thus an increase in interdisciplinary research. Behavioral economics, natural language processing and machine learning have been proved useful in analyzing and modeling the impact of emotional impulses on the behavior of financial markets.

In addition to the development of technology, which resulted in changes in the functioning of financial markets, an equally important issue is the development of the market theory which is both a response to changes and an attempt to explain the mechanisms that form current financial markets. The result is the adaptive market hypothesis (Lo, 2005). The adaptive market hypothesis partly based on the assumptions of behavioral economics points out that price formation is influenced not only by market data, but also by how they are perceived by market participants, which can be examined using sentiment analysis.

Sentiment analysis deals with detecting the general mood prevailing in online resources and social media to understand how people think about a given topic (Nassirtoussi et al., 2015). It is mainly based on identifying positive and negative words and processing text to classify its emotional attitude as positive or negative. Sentiment analysis is based on two assumptions. First, some words express emotions. Second, there are words that can cause emotions when they are spoken (Pang et al., 2002). Thus, on the one hand, sentiment analysis indicates the emotional states of the author of the statement, on the other – it also serves to determine the emotional effect that a given statement can cause (Tomanek, 2014).

The results of research on the impact of information from social media cast new light on the problem of prediction of price changes on capital markets (Johnman et al., 2018; Pagolu et al., 2016; Pasupulety et al., 2019).

The main purpose of this research is to examine the impact of sentiment on the direction of stock price changes on the WIG-banking sub-sector index on the Warsaw Stock Exchange. Specifically, this paper uses several classification machine learning techniques to predict the direction of one-day-ahead stock price changes based on sentiment derived from news headlines.

A detailed analysis of the research results presented in the literature review section allowed for formulating the following hypothesis: The sentiment data extracted from news headlines allow for stock price predictions on the WIG-banking sub-sector index in a one-day time horizon.

It is worth noting that the studies conducted so far have been based, in most cases, on the American market, which has different characteristics compared to the Polish market, i.e. market capitalization, trading volume, number of the investors, specificity of the language, as well as some cultural differences that can cause different perception of the information published. My study attempts to close this gap by implementation of the existing research methods on the Warsaw Stock Exchange.

The paper is organized as follows. Section 2 presents the theory of adaptive markets hypothesis emphasizing the impact of emotional overtones accompanying emerging information about a given entity on the prices of financial instruments. The second part of the chapter presents the issues of sentiment in the context of the possibility of explaining changes on the financial market. Section 3 describes the data used in the study and the methodology of the research. Section 4 presents the results of the study. Finally, the last section concludes.

2. LITERATURE REVIEW

2.1. Adaptive Market Hypothesis

A major part of modern investment theory and practice is based on the Efficient Market Hypothesis (Fama, 1965). This concept assumes that markets fully, accurately and immediately incorporate all available information into market prices. At the root of this far-reaching idea is the assumption that market participants are rational economic entities, always acting in their own interests and making decisions in an optimal way (Lo, 2005). This means that stock prices cannot be predicted because they depend on new information rather than current/past prices. As a result,

stock prices are subject to random walks. In the updated study (Malkiel & Fama, 1970), the authors stated that efficiency can take three forms: weak, semi-strong and strong.

The implication of the Efficient Market Hypothesis is that the market cannot be beaten because all information that could predict performance is already incorporated into the stock price. However, several studies provide evidences contrary to the suggestion of the Efficient Market Hypothesis and Random Walk Theory (Bollen et al., 2011; Schumaker et al., 2012). These studies show that the stock market can be predicted to some extent, and thus question the basic assumptions of the above hypothesis. This phenomenon was explained by behavioral economics, which argues that markets are not efficient, and the element of random walk can be explained by human behavior, because ultimately people are responsible for making decisions, and as people they make irrational and systematic mistakes. These errors affect prices and returns, resulting in inefficiency of the market.

Lo (2005), based on the analysis of the Adaptive Markets Hypothesis (Lo, 2004) attempted to reconcile the Efficient Markets Hypothesis and behavioral economics theory. This hypothesis is based on some well-known principles of evolutionary biology – competition, mutation, reproduction and natural selection. Translating this into the realities of financial markets, this means that the degree of market efficiency is related to environmental factors, such as the number of competitors on the market, the scale of available profit opportunities and the ability to adapt participants to the changing market situation. In other words, it is unrealistic to expect perfectly efficient/inefficient markets – due to behavioral bias. The importance of Adaptive Markets Hypothesis is well documented in the literature (Charles et al., 2012; Hiremath & Narayan, 2016).

Literature provides many examples proving the assumption of Adaptive Markets Hypothesis. Zhou and Lee (2013) analyzed REITs listed on NYSE, AMEX and NASDAQ. Based on the sample of 7,570 daily observations from the period of January 1980 – December 2009, they proved that market efficiency depends on the behavior on given market and is variable over time. Therefore, it cannot be considered as a binary variable, which confirms the assumptions of the Adaptive Market Hypothesis. Kim et al. (2011) evaluated return predictability of the daily and weekly Dow-Jones Industrial Average indices from 1900 to 2009. Based on the analysis, they found strong evidence that stock returns, e.g. during fundamental economic or political crises, have been highly predictable with a moderate degree of uncertainty, which confirms that predictability is driven by changing market conditions.

As a result, building a predictive model based solely on the analysis of historical time series or micro/macroeconomic data puts a big question mark on its effectiveness. The reason for that should be seen in the fact that not only the above-mentioned data have an impact on financial markets, but also on the way they are perceived by market participants. As numerous studies show, sentiment analysis is a factor that significantly improves the effectiveness of prediction.

2.2. Sentiment Analysis

Sentiment analysis deals with detecting the general mood prevailing in online resources and social media to understand how people think about a given topic (Nassirtoussi et al., 2015). It is mainly based on identifying positive and negative words and processing text to classify its emotional attitude as positive or negative. Sentiment analysis is based on two assumptions. First, some words express emotions. Second, there are words that can trigger termination of emotions (Pang et al., 2002).

Thus, on the one hand, sentiment analysis indicates the emotional states of the author, on the other it also serves to determine the emotional effect that a given statement can cause (Tomanek, 2014).

Sentiment analysis is performed at three levels. At the first, the document level, the entire content of the document is classified to determine whether it contains a positive or negative attitude. At the second level, the sentence level, it is determined whether the sentence contains a positive, negative or neutral opinion. Neutral overtones can also mean no opinion. The last level is the entity and aspect level (Liu, 2012).

In sentiment analysis, two methods are used to classify the text:

- dictionary,
- statistical.

The dictionary method can be based on a set of opinion words or the entire corpus of texts. It assumes that there are certain words that are often used to express emotions (Pang et al., 2002). The dictionary method takes into account the meaning of the analyzed words and lexical rules of a given language. Therefore, it is necessary to know the grammatical rules of the analyzed language and the specificity of utterances related to the vocabulary used (Tomanek, 2014).

Second, the statistical method, treats the text as an object, which is represented using quantitative data in the form of, e.g., the number of words or phrases. The statistical method represents the object in the form of a vector in a multi-dimensional space defined by a set of features (Tomanek, 2014).

Sentiment analysis of news can be an effective source for market forecasts, because it expresses the point of view and the mood of opinion leaders who, to some extent, form public opinion and cause public reactions. It is not surprising, then, that the impact of the sentiment of emerging information on price formation has become the focus of extensive research (Hagenau et al., 2013; P. Mehta et al., 2021; Schumaker et al., 2012; Urlam, 2021; Valle-Cruz et al., 2021).

Textual input data used for the sentiment analysis model can have multiple sources. Most of the research has used information gathered from platforms such as Bloomberg (Chatrath et al., 2014; Gumus & Sakar, 2021; Jin et al., 2013) and Yahoo Finance (Rechenthin et al., 2013; Xie et al., 2013).

Periodic financial reports published by the companies are another category of input data sources. It is worth noting that this type of data has a special feature which is the periodicity of publication. According to Huang et al. (2010), a strictly fixed data release schedule may affect predictive capabilities resulting from existing investor expectations for achieving specific financial results. Recently, there has been an increase in interest in the third, less formal source of information such as blogs, microblogs and forums (Yu et al., 2013).

The second category of input data necessary to quantify the impact of sentiment in the context of financial markets are market data in the form of historical quotations for a given financial instrument. Depending on the type of instrument and availability, it may be OHLC data (open-high-low-close) (Y. Mehta et al., 2021; Pagolu et al., 2016; Xie et al., 2013) or only closing price information (Chatrath et al., 2014; Jin et al., 2013; Kumar et al., n.d.). The frequency of the data depends on the frequency of the data containing sentiment.

3. METHODOLOGY AND DATA

3.1. Sample

The objective of the research was to evaluate sentiment derived from news headlines for stock price predictions on the WIG-banking sub-sector index. The composition of the WIG-banking index in the analyzed period, i.e. in the years 2015–2018 is presented in Table 1.

Table1

Composition of the WIB-banking index

Issuer	Ticker	ISIN
Alior Bank SA	ALR	PLALIOR00045
Banco Santander SA	SAN	ES0113900J37
Bank Handlowy w Warszawie SA	BHW	PLBH00000012
Bank Millennium SA	MIL	PLBIG00000016
Bank Ochrony Środowiska SA	BOS	PLBOS00000019
Bank Polska Kasa Opieki SA	PEO	PLPEKAO00016
Getin Holding SA	GTN	PLGSPR000014
Getin Noble Bank SA	GNB	PLGETBK00012
Idea Bank SA	IDA	PLIDEAB00013
ING Bank Śląski SA	ING	PLBSK00000017
mBank SA	MBK	PLBRE00000012
Powszechna Kasa Oszczędności Bank Polski SA	PKO	PLPKO00000016
Bank Zachodni WBK	SPL	PLBZ00000044
UniCredit S.p.A.	UCG	IT0005239360

Source: Self-preparation based on the Stock Exchange Annals published by GPW S.A.

The decision to focus on the WIB-banking index only was taken based on the analysis which showed a dominant share in WIG index across the research period (27.95%, 27.26%, 28.85%, 28.53% respectively). This fact implies the highest number of textual data available for the study among all the indexes. Furthermore, in the analyzed period, the composition of the index did not change, which eliminates potential disturbances resulting from the change in the characteristics of the index.

The study used two data sources: textual data from the InfoStrefa¹ website owned by the Polish Press Agency and the Warsaw Stock Exchange and market data in the form of historical prices of companies in the research sample. The research period is 01.01.2015–31.12.2018 representing a total of 1000 session days. Textual data were extracted based on web scrapping techniques. In order to do that, the web crawler was created.

Historical daily time series was derived from the InfoStrefa. For each session, the following data were collected:

- Open price,
- Minimum and maximum price in each trading session,
- Close price,
- Trading volume,
- Turnover value,
- Number of transactions.

¹ <http://infostrefa.com/infostrefa/pl>. The decision was taken for analysis headers based on the work by Huang et al. (2010), which demonstrated that the headlines are more direct than the entire text – a consequence of a lower level of information noise.

Based on the close price, the daily rate of return was computed in the following way:

$$r_i = \ln\left(\frac{p_i}{p_{i-1}}\right)$$

where:

r_i – daily rate of return,

p_i – stock price in day i ,

p_{i-1} – stock price in day $i-1$.

In the next step, each daily rate of return was labeled in the following way:

$$DIRECTION: \begin{cases} r_i > 0 : 1 \\ r_i \leq 0 : 0 \end{cases}$$

which allows to construct dependent variable “DIRECTION”.

3.2. Data Pre-Processing

Text data were the subject of a few pre-processing steps which contain:

- Tokenization,
- Stop words removal,
- Non-alphanumeric characters removal,
- Conversion text to lower case.

The above steps were performed using the RE and NLTK libraries available for Python.

In the next step the stemming procedure was applied. Stemming was performed based on PoliMorf². Before PoliMorf was applied, the Bug-of-word method (Hájek, 2018) was used to represent the corpus in the form of a sparse matrix.

Extracting the stems from each token made it possible to reduce the number of inflectional forms and thus facilitated the further process of assigning sentiment to each of the headers.

Sentiment assignment was done based on the Polish sentiment dictionary³. The dictionary is a list of words with negative, neutral and positive sentiment. For each of the headline, the number of occurring words with given sentiment was counted. Then, based on the comparison, the overall sentiment of each headline was determined, i.e. if the number of words with negative sentiment prevailed in a given message, then a negative sentiment was assigned. Sentiment was labeled in the following way:

$$sentiment = \begin{cases} -1: negative \\ 0: neutral \\ 1: positive \end{cases}$$

The final stage of data pre-processing was the merging of datasets containing text data with the corresponding market data. The date was used as a key which enabled the merging process. The final dataset included 7,074 observations, of which 3,390 with positive sentiment, 2,665 neutral, and 1,019 negative.

² PoliMorf is the morphological dictionary for Polish resulting from the standardization and merger of Morfeusz SGJP and Morfologik financed by CESAR project.

³ Polish sentiment dictionary was created by The Linguistic Engineering (LE) Group, which is part of the Department of Artificial Intelligence at the Institute of Computer Science, Polish Academy of Sciences.

3.3. Machine Learning

To examine sentiment derived from news headlines for stock price predictions, several machine learning models were applied. Based on the literature review (Jabreel & Moreno, 2018; John & Vechtomova, 2017; Lango et al., 2016), the following algorithms were chosen:

- Gaussian Naive Bayes Classifier,
- Support Vector Classifier,
- KNN Classifier,
- Decision Tree Classifier,
- Random Forest Classifier,
- XGBoost Classifier.

The model training phase was preceded by a procedure of eliminating distortions in both the learning process and the result itself. The procedure consisted of the following:

- verification of data completeness,
- verification of correlations between variables,
- selection of independent variables and the dependent variable,
- splitting the data into training and test set,
- standardization of independent variables.

Data pre-processing as well as machine learning were performed using the Python programming language with dedicated libraries such as Pandas, Numpy, Matplotlib, Seaborn, and Scikit-learn.

Verification of completeness of data was aimed at checking whether there were any variables with missing values in the dataset.

To verify the correlation between variables⁴, the Pearson correlation coefficient was calculated for each pair. As the Pearson correlation coefficient assumes the assumption of linear dependence of variables and normal distribution, for additional verification the Spearman correlation coefficient was also calculated.

In addition, the significance of correlations between strongly correlated variables was examined. A two-sided 95% confidence interval was used for the analysis.

Based on the above analysis, the following variables were selected. Independent variables (x): CLOSE, CHANGE, VOLUME, NUMBER OF TRANSACTIONS, SENTIMENT, COUNT_WORD, MEAN_WORD_LEN; dependent variable (y): DIRECTION.

In accordance with accepted practice found in the literature, the data were split into a training set and a test set based on which the learning performance was assessed. The ratio between the two sets is 80/20.

The last stage preceding model training was the standardization of independent variables.

3.4. Model Evaluation

The predictive power of each model was assessed by comparing the model result with the set “y_test” containing the set of expected values of the DIRECTION variable. For each model, a classification report was prepared. To evaluate the model prediction, two metrics were taken into consideration: accuracy and AUC (Area Under Curve) (Huang & Ling, 2003; Nassirtoussi et al., 2015; Rokach & Maimon, 2014)

In addition, for each model the learning curve was constructed based on 10-fold cross validation to assess if the model is not underfitted/overfitted (Cawley & Talbot, 2010; Guyon, 2009; Guyon et al., 2010).

⁴ The list of all variables with their description is presented in Appendix 1.

Finally, the feature importance was computed, which allowed to increase the interpretability of the model results. The analysis was performed using the permutation importance method available in Eli5 library. The procedure is as follows:

- (1) Get a trained model.
- (2) Shuffle the values in a single column, make predictions using the resulting dataset. Use these predictions and the true target values to calculate how much the loss function suffered from shuffling. That performance deterioration measures the importance of just shuffled variable.
- (3) Return the data to the original order (undoing the shuffle from step 2). Now repeat step 2 with the next column in the dataset until obtaining calculations of the importance of each column.

4. RESULTS

Table 2 presents the aggregated results of all machine learning models.

Table 2
Summary results

	SVC	KNN	Random Forest	Decision Tree	XGBoost	Gaussian Naive Bayes Classifier
Accuracy	0.9611	0.9307	0.9986	0.9957	0.9978	0.9102
AUC	0.9611	0.9307	0.9986	0.9957	0.9978	0.9103

Source: Self-preparation based on empirical results.

Empirical results showed that each model achieved the AUC and accuracy score above 90%. The best classification performance was achieved by CART algorithms, i.e. Decision Tree, Random Forest and XGBoost, where Random Forest has the highest accuracy as well as AUC. The difference between the best CART algorithm, i.e. Random Forest and model with the lowest score – Gaussian Naive Bayes Classifier was 8.84 p.p. for accuracy and 8.83 p.p. for AUC.

Neither of the algorithms showed any problem with underfitting or overfitting. In the case of the KNN Classifier and SVC algorithms, in the initial phase, there was a slight mismatch to the data, which, however, decreased with the increase in the number of training samples.

However, based on the permutation importance technique, it was identified that the impact of the variable SENTIMENT turned out to have an insignificant impact on the prediction result of the algorithms. In other words, the study showed that the sentiment of the data extracted from news headlines does not allow for stock price predictions on the WIG-banking sub-sector index in a one-day time horizon. Therefore, the research hypothesis cannot be accepted. Detailed results of each model including performance metrics, learning curve and future important analysis are presented in the following sections.

4.1. Estimating Results: Decision Tree Classifier

As shown in Figure 1, accuracy and AUC of Decision Tree Classifier is respectively: 0.9958 and 0.9957.

Figure 1

Classification report for Decision Tree Classifier

```

Model Performance metrics:
-----
Accuracy: 0.9958
Precision: 0.9958
Recall: 0.9958
F1 Score: 0.9958

Model Classification report:
-----
              precision    recall  f1-score   support

         1         0.99      1.00      1.00       706
         0         1.00      0.99      1.00       709

   accuracy                1.00      1.00      1.00      1415
  macro avg                1.00      1.00      1.00      1415
 weighted avg                1.00      1.00      1.00      1415

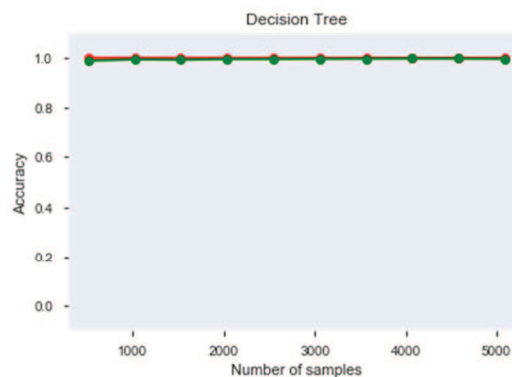
Prediction Confusion Matrix:
-----
      Predicted:
      1      0
Actual: 1      704      2
       0         4     705

AUC: 0.995762694934013

```

Figure 2

Learning curve for Decision Tree Classifier



The learning curve does not show any problem with underfitting or overfitting.

Figure 3

Feature importance report for Decision Tree Classifier

Weight	Feature
0.4779 ± 0.0325	Change
0.0174 ± 0.0038	Number of transactions
0.0113 ± 0.0025	Close
0.0010 ± 0.0014	count_word
0 ± 0.0000	Volume
-0.0001 ± 0.0016	mean_word_len
-0.0006 ± 0.0011	Sentiment

Figure 3 shows the results of feature importance assessment. The values towards the top are the most important features, and those towards the bottom matter least.

The first number in each row shows how much model performance decreased with a random shuffling (in this case, using “accuracy” as the performance metric).

Since there is some randomness in the exact performance change resulting from the shuffling of a specific column, the amount of randomness in permutation importance calculation is computed by repeating the process with multiple shuffles. The number after the \pm measures how performance varied from one reshuffling to the next. The results indicate that variable CHANGE has the most significant impact on model prediction.

4.2. Estimating Results: Random Forest Classifier

Figure 4

Classification report for Random Forest Classifier

```

Model Performance metrics:
-----
Accuracy: 0.9986
Precision: 0.9986
Recall: 0.9986
F1 Score: 0.9986

Model Classification report:
-----
              precision    recall  f1-score   support

         1              1.00      1.00      1.00        706
         0              1.00      1.00      1.00        709

   accuracy                1.00        1415
  macro avg              1.00      1.00      1.00        1415
 weighted avg              1.00      1.00      1.00        1415

Prediction Confusion Matrix:
-----
      Predicted:
      1      0
Actual: 1      705      1
       0       1      708

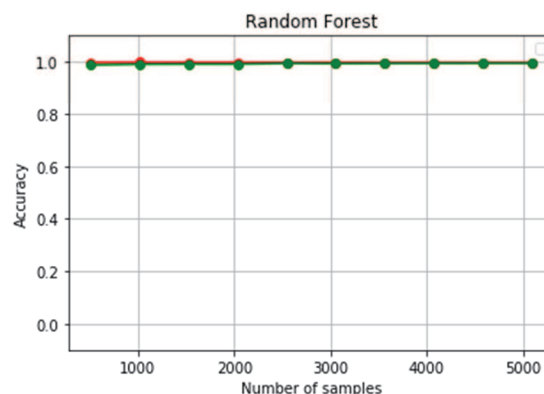
AUC: 0.998586566084778

```

Random Forest Classifier results are slightly better than Decision Tree Classifier (28 bp for both accuracy and AUC).

Figure 5

Learning curve for Random Forest Classifier



Since the shape of learning curve is very close to the Decision Tree Classifier curve, there is no problem with model stability.

Figure 6

Feature importance report for Random Forest Classifier

Weight	Feature
0.4817 ± 0.0311	Change
0.0126 ± 0.0030	Close
0.0105 ± 0.0011	Number of transactions
0.0031 ± 0.0007	Volume
0.0001 ± 0.0006	mean_word_len
0.0001 ± 0.0011	Sentiment
-0.0006 ± 0.0006	count_word

The feature importance report presented in Figure 6 shows that variable CHANGE has the greatest impact on model performance.

4.3. Estimating Results: XGBoost Classifier

Figure 7

Classification report for XGBoost Classifier

```

Model Performance metrics:
-----
Accuracy: 0.9979
Precision: 0.9979
Recall: 0.9979
F1 Score: 0.9979

Model Classification report:
-----
              precision    recall  f1-score   support

         1         1.00      1.00      1.00        706
         0         1.00      1.00      1.00        709

   accuracy                1.00            1415
  macro avg              1.00            1415
 weighted avg              1.00            1415

Prediction Confusion Matrix:
-----
      Predicted:
      1      0
Actual: 1      704      2
       0       1      708

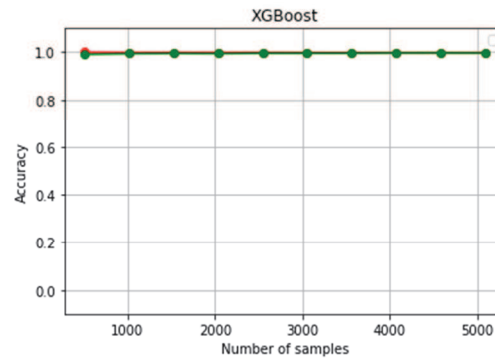
AUC: 0.9978783507873276

```

Performance metrics presented in Figure 7 show that compared to the previous two algorithms, XGBoost ranks between the Decision Tree Classifier and Random Forest Classifier algorithms.

Figure 8

Learning curve for XGBoost Classifier



The learning curve is very similar to the previous two algorithms. Thus, model was not affected by underfitting and overfitting.

Figure 9

Feature importance report for XGBoost Classifier

Weight	Feature
0.4841 ± 0.0293	Change
0.0131 ± 0.0034	Close
0.0045 ± 0.0014	Number of transactions
0.0023 ± 0.0006	Volume
0.0008 ± 0.0011	mean_word_len
0 ± 0.0000	Sentiment
-0.0003 ± 0.0011	count_word

Variable SENTIMENT has no significant impact on model performance, while variable CHANGE was the one with the highest significance.

4.4. Estimating Results: SVC

Figure 10

Classification report for SVC

```

Model Performance metrics:
-----
Accuracy: 0.9611
Precision: 0.9612
Recall: 0.9611
F1 Score: 0.9611

Model Classification report:
-----
              precision    recall  f1-score   support

         1         0.96      0.96      0.96         706
         0         0.96      0.96      0.96         709

   accuracy                0.96         1415
  macro avg              0.96      0.96      0.96         1415
 weighted avg              0.96      0.96      0.96         1415

Prediction Confusion Matrix:
-----
              Predicted:
                1      0
Actual: 1      681    25
         0       30   679

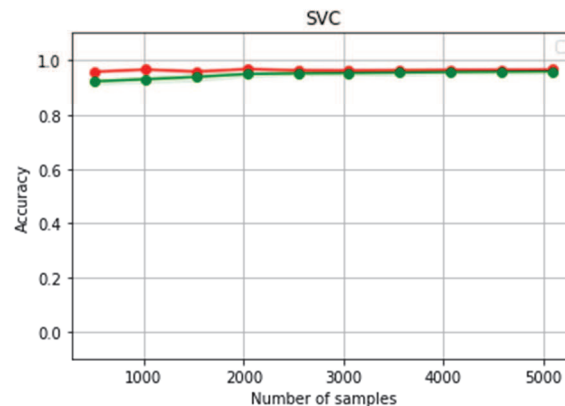
AUC: 0.961138059030594

```

The classification report shows that SVC achieved slightly worse results than CART-based models, i.e. Decision Tree Classifier, Random Forest Classifier and XGBoost Classifier.

The analysis of the learning curve presented in Figure 11 does not indicate model instability.

Figure 11
Learning curve for SVC



The feature importance results presented in Figure 12 are aligned with CART-based models, i.e. showing that variable SENTIMENT has no significant impact on model performance.

Figure 12
Feature importance report for SVC

Weight	Feature
0.4500 ± 0.0262	Change
0.0090 ± 0.0044	count_word
0.0083 ± 0.0035	Volume
0.0073 ± 0.0070	Number of transactions
0.0047 ± 0.0007	mean_word_len
0.0023 ± 0.0024	Close
0.0020 ± 0.0026	Sentiment

4.5. Estimating Results: KNN Classifier

Figure 13
Classification report for KNN Classifier

```

Model Performance metrics:
-----
Accuracy: 0.9208
Precision: 0.9209
Recall: 0.9208
F1 Score: 0.9208

Model Classification report:
-----
              precision    recall  f1-score   support

         1            0.93      0.91      0.92         706
         0            0.92      0.93      0.92         709

   accuracy              0.92              0.92         1415
  macro avg              0.92              0.92      0.92         1415
 weighted avg              0.92              0.92      0.92         1415

Prediction Confusion Matrix:
-----
      Predicted:
      1      0
Actual: 1    645   61
        0     51  658

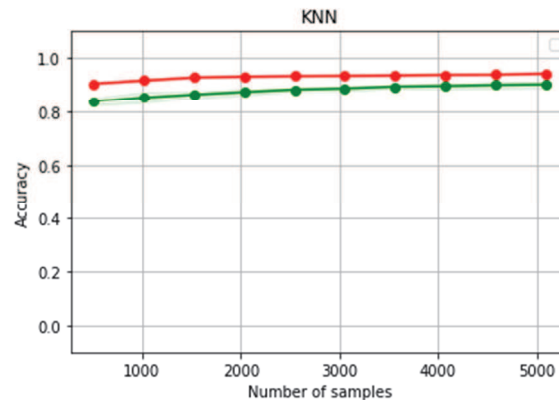
AUC: 0.9307117713573361

```

KNN Classifier performance is definitely lower than CART-based models, e.g. Decision Tree Classifier accuracy was 0.9958 which is better by 7.5 bp.

Figure 14

Learning curve for KNN Classifier



The learning curve presented in Figure 14 indicates a slight variance that decreases as the number of learning samples increases. The increase in the number of learning samples is also accompanied by an increase in classification accuracy.

Figure 15

Feature importance report for KNN Classifier

Weight	Feature
0.3963 ± 0.0267	Change
0.0228 ± 0.0108	mean_word_len
0.0225 ± 0.0067	count_word
0.0212 ± 0.0114	Number of transactions
0.0189 ± 0.0129	Close
0.0133 ± 0.0079	Sentiment
0.0107 ± 0.0076	Volume

The analysis of feature importance is aligned with the results of previous models.

4.6. Estimating Results: Naive Bayes Classifier

The classification report presented in Figure 16 shows that Naive Bayes Classifier performance is the worst among all applied models.

Figure 16

Classification report for Naive Bayes Classifier

```

Model Performance metrics:
-----
Accuracy: 0.9102
Precision: 0.9154
Recall: 0.9102
F1 Score: 0.91

Model Classification report:
-----
              precision    recall  f1-score   support

         1            0.87      0.97      0.91         706
         0            0.96      0.85      0.91         709

   accuracy              0.92
  macro avg              0.92
 weighted avg              0.92

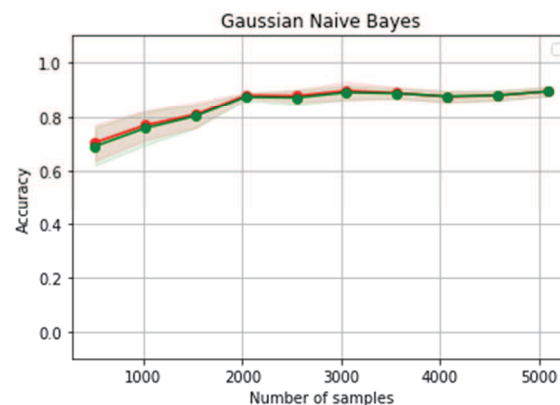
Prediction Confusion Matrix:
-----
              Predicted:
              1      0
Actual: 1      682   24
         0      103  606

AUC: 0.9103653152307243

```

Figure 17

Learning curve for Naive Bayes Classifier



Naive Bayes Classifier was not affected by underfitting and overfitting. It is worth noting the increase in model accuracy with an increase in the number of learning samples up to the level of 2000.

Figure 18

Feature importance report for Naive Bayes Classifier

Weight	Feature
0.3984 ± 0.0247	Change
0.0133 ± 0.0081	Number of transactions
0.0083 ± 0.0027	Close
0.0052 ± 0.0065	mean_word_len
0.0031 ± 0.0029	Volume
0.0007 ± 0.0039	count_word
0.0003 ± 0.0039	Sentiment

Similarly to the other models, analysis of feature importance shows that variable SENTIMENT has no significant impact on model performance.

5. CONCLUSION

The objective of the research was to examine the impact of sentiment derived from news headlines for stock price predictions on the WIG-banking sub-sector index. The text data as well as market data were derived from the InfoStrefa website owned by the Polish Press Agency and the Warsaw Stock Exchange. The research period is 01.01.2015–31.12.2018 representing a total of 1000 session days.

To examine the impact of sentiment, six machine learning models were used: Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, KNN Classifier, SVC and Gaussian Naive Bayes Classifier.

Empirical results show that each model achieved both accuracy and AUC above 90%, i.e. a good ability to predict the direction of price change in one-day time horizon. Furthermore, based on the analysis of learning curves, it was assessed that none of the models was affected by underfitting or overfitting.

However, the results of feature importance analysis show that for each of the model variable SENTIMENT, which contains information about emotional attitude, had no significant impact on its classification performance. Thus, it cannot be concluded that sentiment of news headlines has a significant impact on stock price changes.

Therefore, the conclusion from my study is not similar to the results of most research conducted so far. It is worth noting, however, that the research carried out so far has been based on foreign markets, in most cases on the American market, which has different characteristics compared to the Polish market, i.e. market capitalization, number of investors as well as cultural differences which may cause different perception of published information. In terms of the Polish market, Wojarnik (2021) concluded that sentiment analysis of texts posted on discussion forums may be useful in analyzing the behavior of stock price. However, it should be noted that this research was devoted to three companies from the WIG-GAMES index and a different type of textual data was used.

In addition, it should be taken into account that the financial language, like any other industry language, has a number of specific phrases and terms not used in colloquial speech. According to Loughran and McDonald (2011), who created their own dictionary classifying sentences related to the field of economics and finance, nearly 75% of sentences classified based on the Harvard dictionary as negative after using their dictionary turned out to be positive. The difference was the result of a different sense context.

The PoliMorf and Polish sentiment dictionary were build on the basis of the Polish Language Grammatical Dictionary, which does not take into account the financial context. As a result, there is a risk of imprecisely identified sentiment of messages. Unfortunately, at the time of the research, there was no Polish dictionary available for the financial industry terminology.

On the basis of the conducted research, there are several future directions for this area of research that could be suggested. The first one is to attempt to create a dictionary of financial terms for the Polish language. Perhaps this will lead to better predictivity. The second future direction would be to investigate other machine learning techniques. While classical machine learning models have proven themselves in the textual financial domain, perhaps other more advanced techniques, e.g. deep learning models, could achieve better results. The third future direction would be to explore another source of text data, e.g. social media platforms such as Twitter or Facebook.

References

- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Cawley, G.C., & Talbot, N.L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079–2107.
- Charles, A., Darné, O., & Kim, J.H. (2012). Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates. *Journal of International Money and Finance*, 31, 1607–1626. <https://doi.org/10.1016/j.jimonfin.2012.03.003>
- Chatrath, A., Miao, H., Ramchander, S., & Villupuram, S. (2014). Currency jumps, cojumps and the role of macro news. *Journal of International Money and Finance*, 40, 42–62. <https://doi.org/10.1016/j.jimonfin.2013.08.018>
- Fama, E.F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38, 34–105. <https://doi.org/10.1086/294743>
- Gumus, A., & Sakar, C.O. (2021). Stock market prediction by combining stock price information and sentiment analysis. *International Journal of Advances in Engineering and Pure Sciences*, 33, 18–27.
- Guyon, I. (2009). A practical guide to model selection. In *Proc. Mach. Learn. Summer School Springer Text Stat.* (pp. 1–37).
- Guyon, I., Saffari, A., Dror, G., & Cawley, G. (2010). Model selection: Beyond the bayesian/frequentist divide. *Journal of Machine Learning Research*, 11, 61–87.
- Hagenau, M., Liebmman, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55, 685–697. <https://doi.org/10.1016/j.dss.2013.02.006>
- Hájek, P. (2018). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Computing and Applications*, 29, 343–358. <https://doi.org/10.1007/s00521-017-3194-2>

- Hiremath, G.S., & Narayan, S. (2016). Testing the adaptive market hypothesis and its determinants for the Indian stock markets. *Finance Research Letters*, 19, 173–180. <https://doi.org/10.1016/j.frl.2016.07.009>
- Huang, C.-J., Liao, J.-J., Yang, D.-X., Chang, T.-Y., & Luo, Y.-C. (2010). Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Systems with Applications*, 37, 6409–6413. <https://doi.org/10.1016/j.eswa.2010.02.078>
- Huang, J., & Ling, C.X. (2003). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17, 299–310. <https://doi.org/10.1109/TKDE.2005.50>
- Jabreel, M., & Moreno, A. (2018). EiTAKA at SemEval-2018 Task 1: An ensemble of n-channels ConvNet and XGboost regressors for emotion analysis of tweets. *arXiv preprint arXiv:1802.09233*. <https://doi.org/10.18653/v1/S18-1029>
- Jin, F., Self, N., Saraf, P., Butler, P., Wang, W., & Ramakrishnan, N. (2013). Forex-foreteller: Currency trend modeling using news articles. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1470–1473). ACM. <https://doi.org/10.1145/2487575.2487710>
- John, V., & Vechtomova, O. (2017). Sentiment analysis on financial news headlines using training dataset augmentation. *arXiv preprint arXiv:1707.09448*. <https://doi.org/10.18653/v1/S17-2149>
- Johnman, M., Vanstone, B.J., & Gepp, A. (2018). Predicting FTSE 100 returns and volatility using sentiment analysis. *Accounting & Finance*, 58, 253–274. <https://doi.org/10.1111/acfi.12373>
- Kim, J.H., Shamsuddin, A., & Lim, K.-P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. *Journal of Empirical Finance*, 18, 868–879. <https://doi.org/10.1016/j.jempfin.2011.08.002>
- Kumar, K.S.M.V., Kumar, G.R., & Rao, J.N. (2020) Use sentiment analysis to predict future price movement in the stock market. *International Journal of Advanced Research in Engineering and Technology*, 11, 1123–1130.
- Lango, M., Brzezinski, D., & Stefanowski, J. (2016). PUT at SemEval-2016 Task 4: The ABC of Twitter sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 126–132). <https://doi.org/10.18653/v1/S16-1018>
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5, 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Lo, A.W. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. *Journal of Investment Consulting*, 7, 21–44.
- Lo, A.W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30, 15–29. <https://doi.org/10.3905/jpm.2004.442611>
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66, 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Malkiel, B.G., & Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25, 383–417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
- Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, 7, e476. <https://doi.org/10.7717/peerj-cs.476>
- Mehta, Y., Malhar, A., & Shankarmani, R. (2021). Stock price prediction using machine learning and sentiment analysis. Paper presented at the 2nd International Conference for Emerging Technology (INCET) IEEE. <https://doi.org/10.1109/INCET51464.2021.9456376>
- Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., & Ngo, D.C.L. (2015). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications*, 42, 306–324. <https://doi.org/10.1016/j.eswa.2014.08.004>
- Pagolu, V.S., Reddy, K.N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. Paper presented at the International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs) IEEE. <https://doi.org/10.1109/SCOPEs.2016.7955659>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (Vol. 10, pp. 79–86). Association for Computational Linguistics. <https://doi.org/10.3115/1118693.1118704>
- Pasupulety, U., Anees, A.A., Anmol, S., & Mohan, B.R. (2019). Predicting stock prices using ensemble learning and sentiment analysis. Paper presented at the IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). <https://doi.org/10.1109/AIKE.2019.00045>
- Rechenthin, M., Street, W.N., & Srinivasan, P. (2013). Stock chatter: Using stock sentiment to predict price direction. *Algorithmic Finance*, 2, 169–196. <https://doi.org/10.3233/AF-13025>
- Rokach, L., & Maimon, O. (2014). *Data mining with decision trees: Theory and applications*. World Scientific Publishing Co. <https://doi.org/10.1142/9097>
- Schumaker, R.P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53, 458–464. <https://doi.org/10.1016/j.dss.2012.03.001>

- Tomanek, K. (2014). Analiza sentymentu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych. *Przegląd Socjologii Jakościowej*, 10, 118–136.
- Urlam, S. (2021). Stock market prediction using LSTM and sentiment analysis. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12, 4653–4658.
- Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2021). Does Twitter affect stock market decisions? Financial sentiment analysis during pandemics: A comparative study of the H1N1 and the COVID-19 periods. *Cognitive Computation*, 1–16. <https://doi.org/10.21203/rs.3.rs-39991/v1>
- Wojarnik, G. (2021). Sentiment analysis as a factor included in the forecasts of price changes in the stock exchange. *Procedia Computer Science*, 192, 3176–3183. <https://doi.org/10.1016/j.procs.2021.09.090>
- Xie, B., Passonneau, R., Wu, L., & Creamer, G.G. (2013). Semantic frames to predict stock price movement, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. pp. 873–883.
- Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55, 919–926. <https://doi.org/10.1016/j.dss.2012.12.028>
- Zhou, J., & Lee, J.M. (2013). Adaptive market hypothesis: Evidence from the REIT market. *Applied Financial Economics*, 23, 1649–1662. <https://doi.org/10.1080/09603107.2013.844326>

APPENDIX 1

Description of variables

Variable	Description
COMPANY	Company name
DAY	Date of trading session
TIME	Time of publication of the news
HEADLINE	Headline content
OPEN	Opening price
MAX_PRICE	Highest price
MIN_PRICE	Lowest price
CLOSE	Closed price
CHANGE	The percentage change in the price
DIRECTION	Binary variable: 0 – price decrease/no change, 1 – price increase
VOLUME	Volume
NUMBER OF TRANSACTIONS	Number of executed transactions
TURNOVER	Turnover (in PLN).
SENTIMENT	Variable which contains information about emotional attitude: –1 negative, 0 neutral, 1 positive
COUNT_WORD	Number of words in each headline
COUNT_LETTERS	Number of letters in each headline
MEAN_WORD_LEN	Average length of word in each headline