

Predicting Abnormal Stock Return Volatility using Textual Analysis of News – A Meta-Learning Approach

Renáta Myšková, Petr Hájek, Vladimír Olej
Faculty of Economics and Administration, University of Pardubice,
Studentská 84, 532 10 Pardubice, Czech Republic
renata.myskova@upce.cz, petr.hajek@upce.cz, vladimir.olej@upce.cz

Abstract

Textual analysis of news articles is increasingly important in predicting stock prices. Previous research has intensively utilized the textual analysis of news and other firm-related documents in volatility prediction models. It has been demonstrated that the news may be related to abnormal stock price behavior subsequent to their dissemination. However, previous studies to date have tended to focus on linear regression methods in predicting volatility. Here, we show that non-linear models can be effectively employed to explain the residual variance of the stock price. Moreover, we use meta-learning approach to simulate the decision-making process of various investors. The results suggest that this approach significantly improves the prediction accuracy of abnormal stock return volatility. The fact that the length of news articles is more important than news sentiment in predicting stock return volatility is another important finding. Notably, we show that Rotation forest performs particularly well in terms of both the accuracy of abnormal stock return volatility and the performance on imbalanced volatility data.

Keywords: stock return volatility, prediction, textual analysis, sentiment, meta-learning

JEL Classification: G12, G17, C45, C53

Introduction

Key investment decisions are based on the assumption that the theoretical stock price, expressed by its intrinsic value, is different from the current stock market price for which it is traded on organized markets. The intrinsic value of the stock can be considered constant in a very short period of time because investor's subjective opinion is not expected to change rapidly. However, each investor compares the perceived intrinsic value with its market value and makes buy/hold/sell decisions based on the current situation. Technical analysis of stocks is based on the following theoretical assumptions: (1) investors' behavior is basically predictable (investors' reactions in stock markets repeat), and (2) stock prices are formed on the basis of market supply and demand and reflect not only the available data from the fundamental analysis but also the optimism or pessimism of the stock markets' participants. Investors' expectations, including the technical analysis of the stocks and their subjective reactions, thus form market stock prices. Changes in the prices are therefore also related to psychological analysis, based on the assumption that stock prices are strongly influenced by the psychological response of investors, especially in the short-term.

What is crucial from the point of view of firm-related text news: when investors' expectations change, the intrinsic values of stocks change in their mindset too and, as a result, their willingness to buy or sell stocks is affected. This is reflected in the stock price. Drasnar (1995) explains this process in his concept of psychological analysis as the effect of contradictory human characteristics - greed and fear. Depending on what currently prevails and how rapidly the characteristics change, increasing and decreasing trends alternate. In fact, recent studies suggest that public information diffuses gradually through the investor population (DellaVigna and Pollet, 2009).

Changes in stock prices, as well as changes in the stock market as such, can be expressed in terms of volatility. In this paper, the impact of news articles published at Yahoo Finance on the abnormal stock return change is incorporated in prediction models. This effect is expected as the result of psychological impact on existing and potential investors. Indeed, the effect of media news is increasingly recognized as key determinant in financial markets. For example, Peress (2014) reported that the volatility of stock returns is significantly affected by news, suggesting that the media contribute to the efficiency of the stock market (information is disseminated among investors and incorporated into stock prices).

Several studies have recently demonstrated that the textual analysis of news and other firm-related documents may improve the accuracy of volatility prediction models (Tetlock, 2007; Kothari et al., 2009; Loughran and McDonald, 2011). However, the research to date has tended to focus on econometrics and statistical prediction methods rather than machine learning despite the fact that methods like Support Vector Machines (SVMs) (Groth and Muntermann, 2011) outperformed traditional statistical methods in volatility prediction in terms of accuracy. This may be explained by the fact that non-linear machine learning models can be employed to explain the residual variance of the stock price from the prediction of the regression equation (Tsai et al., 2011). Thus, better prediction performance can be achieved. Moreover, recent empirical evidence suggests that meta-learning (ensemble) approaches can be more effective in predicting stock market movements than single machine learning methods (Tsai et al., 2011; Zhou et al., 2012; Hajek et al., 2015; Ballings et al., 2015). This evidence results from the fact that the combination of single prediction models is able to complement the errors made by the individual models on different parts of the data space. The performance of meta-learning models is likely better than that of the best single classifiers used in isolation (Tsai et al., 2011). Therefore, here we develop a novel meta-learning model based on textual analysis of news to predict abnormal stock return volatility. This model is also based on the assumption that not all investors react to news in the same way. We hypothesize that meta-learning approaches can effectively simulate the decision-making process of several types of investors and thus improve the prediction accuracy.

The remainder of this paper is organized as follows. Section 2 presents the related literature on volatility prediction using textual analysis. In the next section, the theoretical foundations of stock return volatility (and stock market volatility) are introduced. Section 4 describes data collection and their processing. Section 4 presents the results of experiments and compares the performance of meta-learning models with single prediction models. Our conclusions are drawn in the final section.

Related Literature

Previous literature has examined the effects of several text sources on stock market prices. Corporate annual reports have been used in several studies related to predicting overall firm financial performance (Hajek et al., 2014), abnormal stock returns (Hajek and Bohacova, 2016) or market-to-book ratio (Myskova and Hajek, 2016). The importance of information contained in the firm's press releases and quarterly earnings announcements was studied by Demers and Vega (2010), Davis et al. (2012), and Ahern and Sosyura (2014). Price et al. (2012) also examined the impact of data presented at press conferences. Clatworthy and Jones (2003) evaluated which reports are presented by top management in relation to pre-tax profits. Their research showed that there is a tendency to choose an approach that promotes management's interests, this is positive (good) news are primarily related to the firm, whereas negative (bad) news are associated with external influences. This corresponds to the findings that positive news increase stock price volatility (Kothari and Short, 2009) and predict a favorable future development (Li, 2010).

News in the information media are particularly important if the firm does not present itself too much. In that case, investors do not have enough explanatory commentary for both retrospective and current analysis and, thus, any news may be given unjustified importance. According to Huang et al. (2014), the announcement of positive economic results in press releases has led to a sharp increase in returns on the stock market following such announcements. This finding also corresponds to Barber and Odean (2008), Da et al. (2011), and Davis et al. (2012). DellaVigna and Pollet (2009) and Fang and Peress (2009) mention the need for press releases to reduce their asymmetry of information. Meluzin and Zinecker (2014, 2016) investigated the role of the media in IPOs and concluded that firms with larger media support have higher initial returns and increase their value over the long term. Using social media to transmit positive signals towards investors affect the IPO success significantly.

Table 1 presents the list of previous studies on volatility prediction based on textual analysis.

Table 1: List of prior studies on volatility prediction using textual analysis

Study	Textual source	Features linguistic / financial	Method	Key findings
Antweiler and Frank (2004)	Message boards (Yahoo!Finance) News (Wall Street Journal)	No. of messages, bullishness index, agreement index / stock market index, trading volume	GARCH	More messages today imply greater market volatility tomorrow
Tetlock (2007)	News (Wall Street Journal)	General Inquirer (pessimism, negative, weak word categories) / past volatility, trading volume	OLS	Pessimism weakly predicts increases in market volatility
Kothari et al (2009)	News (Dow Jones Interactive and Factiva), corporate reports (EDGAR) and analyst disclosure (Investext and Factiva)	General Inquirer (favorable, unfavorable) / market capitalization, book-to-market equity, leverage ratio	Fama-MacBeth regression	Favorable (unfavorable) with highly significant negative (positive) effect on return volatility
Loughran and McDonald (2011)	corporate reports (EDGAR)	Positive, negative, uncertainty, litigious, modal / market capitalization, book-to-market equity, share turnover, institutional ownership, Fama-French alpha	Fama-MacBeth regression	positive, negative or modal words positively affect stock return volatility in the year after the corporate report filing
Groth and Muntermann (2011)	News data from a collection of corporate disclosures	<i>tf-idf</i> of words selected using Chi-Squared-based feature selection	NB, k-NN, MLP, SVM	textual data represents a valuable source of information for financial risk management, SVM method performed best in terms of accuracy and AUC

Legend: OLS is ordinary least square, *tf-idf* is term frequency-inverse document frequency, NB is Naive Bayes, k-NN is k nearest neighbors, SVM is support vector machine, AUC is the area under the receiver operating characteristic curve.

Antweiler and Frank (2004) combined texts from message boards and news to demonstrate that more frequent messages/news imply greater volatility in future. In addition to information extracted from the text documents, the authors used stock market index and trading volume as the financial determinants of market volatility. Tetlock (2007) also used the news provided by the Wall Street Journal to study the effect of news' sentiment on market volatility. Pessimism in the news seems to be effective in market volatility prediction. Kothari et al. (2009) combined

three text resources to show that favourable and unfavourable news have significant effect on stock return volatility. A long-term effect of sentiment in annual reports on stock return volatility was observed by Loughran and McDonald (2011). Specific financial dictionaries were developed to measure various categories of sentiment in the text. Finally, Groth and Muntermann (2011) first extracted the frequencies of words from the corpus of news and then used machine learning to accurately predict abnormal stock return volatility.

Stock Volatility

Generally, volatility can be considered as a measure describing the rate of fluctuation of the asset's value or its return over a period of time. It is usually defined as the standard deviation of the asset price. This statistical measure of price variance over a given period determines the risk of the investment instrument, as generally the higher the volatility of the instrument, the higher the risk of the investment instrument (its price is more flexible). The standard deviation of the stock return can be calculated as follows:

$$\sigma_t(n) = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (R_t - \bar{R})^2}, \quad (1)$$

where R_t is the stock return at time t , \bar{R} denotes the average stock price return for the number n of periods, and the period n days depends generally on the length of the investment strategy. The relative change of the closing price of a security to its value before the specified trading days can be expressed by the stock price return R (also known as the Rate of Change):

$$R_t(n) = \frac{P_t - P_{t-n}}{P_{t-n}} * 100. \quad (2)$$

In this study, one-day return was used, this is $R_t(1) = (P_t - P_{t-1})/P_{t-1}$, where P_t is the closing price at time t .

Based on other criteria, normal volatility, historical volatility, correlated volatility or implied volatility can be distinguished. From the investors' point of view, capital market volatility may also be important. This can be used as a benchmark in assessing the volatility of a particular asset. Using the Beta coefficient, it is possible to compare the volatility of the asset with market developments, this is stock volatility against the relevant stock index. In this case, it is a correlated volatility (Beta). Higher asset price volatility, which is determined as a comparison of historical prices over a short period with the latest (current) price, indicates an increased risk because its price becomes less predictable. In addition to the so-called "common volatility", the historical volatility of a financial instrument (calculated by using past data) is also often measured. The implied volatility is the volatility of a financial derivative (usually an option), and it is derived from the volatility of the underlying asset.

When deciding on an investment, a high volatility of an asset may significantly affect the interest of the potential investor, as well as the decision of the existing investor to sell the asset. Upon purchase, the high volatility allows investors to buy the asset after a short-term decline and later sell at an overvalued price. However, when selling, high volatility is often reflected in a price decrease and thus in the depreciation of the asset being sold.

According to Charles Dow's theory (Edwards et al., 2007), the evolution of most stocks is in the same direction as the stock market. The behavior of market participants are influenced by their expectations and therefore stock indices reflect these expectations. Consequently, stock price developments need to be assessed not only in the time context of the stock title, but also in relation to the development of the relevant stock market index. The relative volatility, expressed by Beta coefficient, assesses the correlation between the price of a particular stock.

If Beta ranges from 0 to 1 (not including one of the extreme values), the stock volatility is less than that of the stock index, at 1 the volatility is exactly the same, and the values above 1 indicate the more volatile stock than the stock market index.

When evaluating stock volatility, technical indicators of the stock cannot be ignored. These are price indicators, including moving averages, bands and oscillators. Moving averages are most frequently used indicators, but their ability to smooth out price fluctuations does not make them appropriate to analyze news sentiment and evaluate its impact on volatility. For the same reason, we also excluded bands because its principle of creating an envelope representing the band around the moving average curve does not capture volatility changes induced in a short period of time by published news. From the technical indicators, therefore, oscillators are best fitted to measure the change in the stock price over a selected period of time. They can be used on bull (upward) or bear (downward) markets, and the shorter the period of the observed period, the higher their ability to indicate the signals of future changes.

The "momentum" oscillator measures the acceleration or slowdown of the stock price trend by comparing the current closing stock prices with the prices at the beginning of the reference period. It can be used in absolute or relative terms. To monitor price fluctuations at several-day intervals, it is better to use a relative momentum reflecting how much the price has changed over the time period n between the last price value and the value being compared. The value of n can vary between 3 and 20 days. The positive value of the indicator is significant not because it presents an upward trend, but it is important whether it increases or not. The increase indicates a positive trend, while a decreasing trend can indicate future trend correction or even change in its direction. Stochastic oscillator $\%K_t(n)$ is another momentum indicator comparing the closing price to the range of its prices over a specified period of time:

$$\%K_t(n) = 100 * \frac{P_t - P_{low}}{P_{high} - P_{low}}, \quad (3)$$

where P_{low} is the lowest price of the n previous trading days and P_{high} is the highest price traded during the same n -day period.

Data and Research Methodology

Financial Data

To select firms for the prediction of abnormal stock return volatility, we used the following criteria. The first criterion was that the U.S. firms had to be listed on a major stock exchange, specifically the New York Stock Exchange (NYSE) or Nasdaq. Second, to reduce the contribution of bid/ask bounce in reaction to news, we followed Loughran and McDonald (2011) and included only those firms with a reported stock price of at least 3 USD before the news release. To reduce the effect of risk factors for stocks, high market capitalization was required (Fama and French, 1993). Specifically, we chose 14 largest U.S. firms in terms of market capitalization, see Table 2. We downloaded all news for such firms from the Yahoo Finance system (finance.yahoo.com) for the period February 2016. This period was chosen mainly because no significant fluctuations were observed in macroeconomic sentiment (measured by the Dow Jones Economic Sentiment Indicator). Thus, the effect of macroeconomic sentiment on the firms' volatility was reduced. To calculate the indicators of stock return (and price) volatility $\sigma_t(n)$ (and $\%K_t(n)$), we used the period of three days, $n = 3$, where $t-3$ represents the release of news. This three-day event window is consistent with related studies (Loughran and McDonald, 2011). In addition, we also measured the one-day stock return to control its impact on abnormal stock return volatility. The average values of the indicators are presented in Table 2. The results show that ConocoPhillips and Morgan Stanley, respectively, had the highest stock return volatility (in terms of standard deviation), while the oscillator indicator shows that Exxon Mobile Corp. had the highest stock price volatility.

Table 2: Average values of volatility indicators for selected U.S. firms

Firm	$\sigma_t(3)$	$R_{t-3}(1)$	$\%K_{t-3}(3)$
Apple	0.021	0.002	0.421
Bershire Hathaway	0.013	0.006	0.590
Cisco	0.027	0.024	0.559
ConocoPhillips	0.044	-0.021	0.413
Costco	0.015	0.001	0.549
Ebay	0.027	-0.002	0.476
Exxon Mobile Corp.	0.017	0.013	0.656
Ford Motor	0.026	0.009	0.523
General Electric	0.019	0.005	0.528
Morgan Stanley	0.037	-0.009	0.421
Pepsico	0.011	0.004	0.576
The Procter & Gamble Co.	0.013	0.004	0.583
Visa	0.027	0.003	0.590
Wells Fargo	0.024	-0.003	0.486

To control other financial indicators that could affect future stock return volatility, we also measured the logarithms of market capitalization and volume (to control risk factors for stocks) in time $t-3$, past stock return volatility ($\sigma_{t-3}(3)$), and stock market (Nasdaq / NYSE). The descriptive statistics of these predictors are presented in Table 3.

Table 3: Mean and St.Dev. values of remaining financial determinants

Indicator	Mean \pm St.Dev.
log (market capitalization)	2.111 \pm 0.389
log (trading volume)	7.175 \pm 0.392
$\sigma_{t-3}(3)$	0.024 \pm 0.016
Nasdaq / NYSE occurrences	64 / 160

Following previous studies (Groth and Muntermann, 2011), abnormal return volatility was calculated as stock return volatility in excess of the stock market return volatility. If stock return volatility was higher than stock market return volatility, the value of 1 was assigned, otherwise 0 was assigned. Thus, the firms were categorized into two classes, with positive (higher than stock market return volatility with 148 firm-day occurrences) and negative abnormal return volatility (76 firm-day occurrences). This suggests that the data were imbalanced in favor of abnormal volatile firms.

Linguistic Data

All the news collected for each firm-day were merged into one text file. Thus, the results of textual analysis could easily be matched with the financial determinants mentioned above. To analyze the text of news, we compared the text files with two sentiment word categorizations, positive and negative. These word categorizations were developed specifically for financial domain by Loughran and McDonald (2011) and represent the most commonly used sentiment dictionaries used in related financial prediction problems. The authors of these dictionaries created the extensive word lists of 354 positive and 2,329 negative words (available at http://www3.nd.edu/~mcdonald/Word_Lists.html). While the use of negative words was reported to be unambiguous, the issue of using positive words in a negative statement had to be addressed. We adopted the approach proposed by the authors of the dictionaries and performed

a collocation analysis with positive words to detect one of six negation words (no, not, none, neither, never, nobody) occurring within three words preceding a positive word. These occurrences were then subtracted from the total number of positive words.

Raw term frequencies of negative and positive words used in the news at the day in time $t-3$ were used to measure the sentiment of news. This term frequency scheme has been preferred in many related studies (Doran et al., 2012; Garcia, 2013). In this scheme, the length of texts is considered by normalizing the sentiment category counts by the length of the firm-day news as follows:

$$\text{Positive}_{t-3}(1) = \frac{f_{\text{positive}}}{f_{\text{total}}}, \quad (4)$$

$$\text{Negative}_{t-3}(1) = \frac{f_{\text{negative}}}{f_{\text{total}}}, \quad (5)$$

where f_{positive} and f_{negative} are the raw term frequencies of positive and negative words, respectively, and f_{total} is the length of document. To measure the overall (net positive) sentiment, we used the following formula (Henry, 2008):

$$\text{Sentiment}_{t-3}(1) = \frac{\text{Positive}_{t-3}(n) - \text{Negative}_{t-3}(n)}{\text{Positive}_{t-3}(n) + \text{Negative}_{t-3}(n)}. \quad (6)$$

Table 4 shows the descriptive statistics of the linguistic indicators obtained from the text analysis.

Table 4: Mean and St.Dev. values of linguistic volatility determinants for positive and negative classes of abnormal return volatility

Indicator	Positive class	Negative class
$\text{positive}_{t-3}(1)$	0.0018±0.0009	0.0020±0.0009
$\text{negative}_{t-3}(1)$	0.0022±0.0012	0.0023±0.0013
$f_{\text{total } t-3}(1)$	17,275±14,388	16,519±12,316
$\text{sentiment}_{t-3}(1)$	-0.0748±0.3266	-0.0387±0.3304

The results show that positive class (with a high risk – abnormally high return volatility) is associated with a more negative overall sentiment and longer (or more frequent) news. These findings corroborate those reported by Antweiler and Frank (2004) (more frequent news imply higher stock return volatility) and Tetlock (2007) (pessimism implies higher stock return volatility). Moreover, we also support the empirical evidence obtained by Loughran and McDonald (2011) that positive and negative sentiment promotes stock return volatility.

Prediction Models

As a result, the prediction problem can be defined as follows:

$$\sigma_t(3) = f(R_{t-3}(1), \%K_{t-3}(3), \log(\text{MC}_{t-3}) \log(\text{Vol}_{t-3}), \sigma_{t-3}(3), \text{SE_Type}, \text{positive}_{t-3}(1), \text{negative}_{t-3}(1), f_{\text{total } t-3}(1), \text{sentiment}_{t-3}(1)), \quad (7)$$

where $\sigma_t(3)$ was represented by positive/negative class, MC denotes market capitalization, Vol is trading volume, and SE is stock exchange.

Since the task was to classify stock return volatility with a high accuracy, several machine learning methods were employed. Formally, classification into two classes can be defined as follows. Each instance I is mapped to one element of the set $\{P, N\}$ of positive (P) and negative (N) class labels. A classification model (or classifier) is a mapping from instances to predicted

classes. To distinguish between the actual class and the predicted class we use the labels $\{Y, N\}$ for the class predictions produced by a model.

In this study, we used the following meta-learning algorithms for binary classification: (1) Decorate (Melville and Mooney, 2003); (2) Random subspace (Ho, 1998); (3) Adaboosting (Freund and Schapire, 1996); (4) Random committee (Witten and Frank, 2005); (5) Bagging (Breiman, 1996); (6) Multiboosting (Webb, 2000); and (7) Rotation forest (Rodriguez et al., 2006). These algorithms worked over random tree method as a base classifier.

Decorate meta-learning algorithm uses an existing strong learner (one that provides high accuracy on the training data) to build an effective diverse committee by adding different randomly constructed examples to the training set when building new committee members. Random subspace (or attribute bagging) (Ho, 1998; Bryll, 2003) utilizes many base learners which are systematically constructed by pseudorandomly selecting subsets of components of the feature vector. This method constructs a decision tree based classifier that maintains highest accuracy on training data and improves on generalization accuracy as it grows in complexity. Adaboosting (Freund and Schapire, 1996; Freund and Schapire, 1997) is an algorithm for constructing a strong classifier as linear combination. This meta-learning algorithm is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers and is sensitive to noisy data and outliers. However, in some problems it can be less susceptible to the over fitting problem than most learning algorithms. Random committee algorithm builds an ensemble of base classifiers (with different random number seeds) and averages their predictions (Witten and Frank, 2005). Bagging (Breiman, 1994; Breiman, 1996) is a meta-learning algorithm that includes training many classifiers on different partitions of the training data and using the majority vote on the results of all those classifiers to define the final answer for a test pattern. Multiboosting (Webb, 2000) extends the approach of Adaboosting with the wagging technique, which is a variant of bagging but where the training weights generated during boosting are utilized in the selection of the bootstrap samples. Finally, Rotation forest (Rodriguez et al., 2006) draws upon the random forest idea. The base classifiers are also independently built decision trees, but in rotation forest each tree trained on the whole data set in a rotated feature space. As the tree learning algorithm builds the classification regions using hyperplanes parallel to the feature axes, a small rotation of the axes may lead to a very different tree.

Experimental Results

To avoid overfitting, we used 10-fold cross-validation in our experiments, this is 90% as training and 10% as testing data repeated 10 times. The classification performance was measured by the averages (over the 10 experiments) of standard statistics applied in classification tasks (Powers, 2011), see Table 5.

Table 5: Classification performance measures

Abbreviation	Description	Formula
Acc [%]	accuracy	$(TP+TN)/(TP+FP+TN+FN)$
TP rate	true positive rate	$TP/(TP+FN)$
FP rate	true negative rate	$TN/(TN+FP)$
Pre	precision	$TP/(TP+FP)$
Re	recall	$TP/(TP+FN)$
F-m	F-measure	$2 \times (Pre \times Re) / (Pre + Re)$
ROC	receiver operating characteristic	area under ROC curve (combinations of TP and FP rates)

Legend: TP is the number of instances classified as true positive, TN as true negative, FP as false positive and FN as false negative.

The ROC is a graphical plot that illustrates the performance of a binary classifier system. Specifically, it represents a standard technique for the summarization of classifier performance over a range of trade-offs between TP and FP error rates. It is therefore suitable for imbalanced classification problems.

Table 6 shows the best results of the analyzed methods. The best results were obtained after an extensive series of experiments with various model parameters' settings. Specifically, the classification performance of the meta-learning algorithms was tested for the following user-defined parameters:

- 1) Decorate (D): artificial size = (1, 2, ..., 10), no. of iterations = (5, 10, 20, ..., 200)
- 2) Random subspace (RSS): no. of iterations = (5, 10, 20, ..., 250), sub space size = (0.1, 0.2, ..., 0.9)
- 3) Adaboosting (AB): no. of iterations = (5, 10, 20, ..., 200), weight threshold = (10, 20, ..., 100)
- 4) Random committee (RC): no. of iterations = (5, 10, 20, ..., 200)
- 5) Bagging (B): no. of iterations = (5, 10, 20, ..., 200), bag size percent = (5, 10, 20, ..., 100)
- 6) Multiboosting (MB): no. of iterations = (5, 10, 20, ..., 200)
- 7) Rotation forest (RF): max. group = (1, 2, ..., 40), min. group = (1, 2, ..., 40), no. of iterations = (5, 10, 20, ..., 200), removed percentage = (5, 10, 20, ..., 100).

Random tree (RT) as a base classifier in all the meta-learning algorithms was tested for: no. of randomly chosen attributes $K = (1, 2, \dots, 10)$, the maximum depth of the tree = (1, 2, ..., 20), and the minimum total weight of the instances in a leaf = (1.0, 2.0, ..., 10.0). All experiments were carried out in Weka 3.7.

Table 6: Best results of the analyzed meta-learning algorithms

	D		RSS		AB		RC		B		MB		RF	
Acc [%]	78.5714		79.2857		79.2857		79.6429		80.0010		80.3571		81.4286	
Class	0	1	0	1	0	1	0	1	0	1	0	1	0	1
TP rate	0.758	0.811	0.765	0.818	0.803	0.784	0.773	0.818	0.780	0.818	0.780	0.824	0.780	0.845
FP rate	0.189	0.242	0.182	0.235	0.216	0.197	0.182	0.227	0.182	0.220	0.176	0.220	0.155	0.220
Precision	0.781	0.789	0.789	0.796	0.768	0.817	0.791	0.801	0.792	0.807	0.798	0.808	0.817	0.812
Recall	0.758	0.811	0.765	0.818	0.803	0.784	0.773	0.818	0.780	0.818	0.780	0.824	0.780	0.845
F-m	0.769	0.800	0.777	0.807	0.785	0.800	0.782	0.809	0.786	0.812	0.789	0.816	0.798	0.828
ROC	0.852		0.841		0.867		0.837		0.847		0.861		0.876	

Legend: D: artificial size = 2, no. of iterations = 20 (RT: $K = 6$, max. depth = 20, min. num. = 2.0); RSS: no. of iterations = 150, sub space size = 0.5 (RT: $K = 0$, max. depth = 20, min. num. = 1.0); AB: no. of iterations = 16, weight threshold = 100 (RT: $K = 6$, max. depth = 10, min. num. = 2.0); RC: no. of iterations = 58 (RT: $K = 3$, max. depth = 18, min. num. = 8.0); B: bag size percent = 100, no. of iterations = 80 (RT: $K = 5$, max. depth = 20, min. num. = 7.0); MB: no. of iterations = 22 (RT: $K = 0$, max. depth = 10, min. num. = 2.0); RF: max. group = 30, min. group = 20, no. of iterations = 50 (RT: $K = 0$, max. depth = 1, min. num. = 2.0).

The results presented in Table 6 show the performance of each algorithm. The results show that Decorate has the weakest performance. Then follow Random subspace and Adaboosting. They had the same accuracy Acc[%] (but different precision and recall characteristics). Notably, Adaboosting performed well on both classes, resulting in a high ROC value. Acc[%] is not the only metric for evaluating the effectiveness of a classifier. Two other useful metrics are precision and recall. These two metrics can provide much greater insight into the performance characteristic of binary classifiers. Precision measures the exactness of classifiers. A higher precision means less FP, while a lower precision means more FP. This is often at odds with recall, as an easy way to improve precision is to decrease recall. Recall measures the completeness, or sensitivity, of a classifier. Higher recall means less FN, while lower recall means more FN. Improving recall can often decrease precision because it gets increasingly harder to be precision as the sample space increases. This is important due to the fact that

positive abnormal stock return volatility prevailed in the data. Moreover, it is usually more important to correctly classify firms with a high risk (class 1), rather than those with a low risk (class 0). The group of meta-learning algorithms Random committee, Bagging and Multiboosting performed well, especially on class 1 (with a high TP rate). Rotation forest meta-learning algorithm performed best in terms of Acc[%] and ROC. This suggests that this classifier is suitable for this type of imbalanced problems. In fact, this method performed well on both classes, with TP rate = 0.780 (i.e. 78.0 %) for class 0 and TP rate = 0.845 (84.5 %) for class 1. However, note that the prediction accuracy Acc[%] has to be interpreted considering the fact that class 1 predominated in the data with 66.1 %. Each classifier that is better than a random prediction should therefore perform better than this percentage. This is true for all meta-learning approaches.

Principal component analysis (PCA) was used as a projection filter in the RF algorithm. Table 7 shows the sensitivity to the PCA setting. Specifically, the number of principal components to retain was the proportion of variance covered. This parameter was tested for the following values: 0.1, 0.2, ..., 1.0. Table 7 indicates that the RF performed best with the variance covered = 0.5. In addition, a reasonable tree size of 105 leafs was achieved using this setting. This suggests that highly interpretable models can be obtained using meta-learning algorithms.

Table 7: Sensitivity analysis of RF to PCA setting

Variance covered	ROC	Tree size (no. of leafs)
0.1	0.858	125
0.2	0.858	125
0.3	0.858	125
0.4	0.869	125
0.5	0.876	105
0.6	0.870	129
0.7	0.873	129
0.8	0.870	115
0.9	0.870	115
1.0	0.869	115

For comparison with standard machine learning methods used in previous related studies (Groth and Muntermann 2011; Hajek and Olej, 2013), the results are presented in Table 8 also for support vector machine (SVM), RT, and multilayer perceptron (MLP). The results in Table 8 show that these methods do not achieve the performance of the meta-learning algorithms. SVM uses kernel functions to separate the hyperplane between two classes by maximizing the margin between the closest data points. This is done in a higher-dimensional space where the data become linearly separable. We used the SVM trained by the sequential minimal optimization algorithm. The classification performance of the SVMs was tested for the following user-defined parameters: the complexity parameter $C = (1,2,4,8, \dots, 64)$, round-off error epsilon = (1.0E-10,1.0E-12,1.0E-14), random seed = (1,2,3), tolerance parameter = (0.001,0.01,0.1), kernel functions = (puk, linear, RBF). The classification performance of the RT was tested as presented above, while MLP was trained using the backpropagation algorithm with momentum. The following parameters of the MLP were examined: the number of neurons in the hidden layer $a = (\text{no. of input and output variables})/2$, learning rate = (0.1,0.2, ..., 0.5), momentum = (0.1,0.2, ..., 0.5), seed = (1,2, ..., 5), no. of iterations = (5,10,20,100, ..., 800), and validation threshold = (1,2,3).

Table 8: Best results of the analyzed methods

	SVM	RT	MLP

Acc [%]	73.9286		74.2857		78.2143	
Class	0	1	0	1	0	1
TP rate	0.735	0.743	0.785	0.730	0.659	0.892
FP rate	0.257	0.265	0.270	0.242	0.108	0.341
Precision	0.719	0.759	0.714	0.771	0.845	0.746
Recall	0.735	0.743	0.758	0.730	0.659	0.892
F-m	0.727	0.751	0.735	0.750	0.740	0.812
ROC	0.786		0.760		0.795	

Legend: SMO: the complexity parameter $C = 4$, epsilon = $1E-12$, random seed = 2, tolerance parameter = 0.001, the kernel to use = puk with omega and sigma = 1.0; RT: $K = 4$, max. depth = 14, min. num. = 2.0; MLP: hidden layers = a , learning rate = 0.3, momentum = 0.4, seed = 4, no. of iterations = 600, validation threshold = 2.

To test the differences in performance between the best meta-learning model (Rotation forest) and the used standard machine learning methods, we employed Student's paired t -test. We observed statistical significant differences in terms of both the Acc[%] and the ROC at $p = 0.05$. We can therefore conclude that the proposed meta-learning model significantly outperformed previously used standard machine learning methods.

Finally, we tested the merits of all the input variables using the Relief algorithm (Kira and Rendell, 1992), see Table 9. This algorithm evaluates the importance of a variable by repeatedly sampling an instance and considering the value of the given variable for the nearest instance of the same/different class.

Table 9: Ranking of the input variables using the Relief algorithm

Indicator	Merit	Rank
$\log(MC_{t-3})$	0.062±0.003	1.4±0.5
SE_Type	0.059±0.009	1.6±0.5
$\log(Vol_{t-3})$	0.033±0.004	3.0±0.0
$\sigma_{t-3}(3)$	0.015±0.002	4.0±0.0
$f_{total\ t-3}(1)$	0.007±0.001	5.3±0.5
$\%K_{t-3}(3)$	0.003±0.004	6.9±1.7
$R_{t-3}(1)$	0.002±0.001	7.2±0.6
$positive_{t-3}(1)$	0.002±0.003	7.4±1.4
$negative_{t-3}(1)$	-0.001±0.002	8.6±0.9
$sentiment_{t-3}(1)$	-0.002±0.003	9.6±0.5

Legend: MC is market capitalization, Vol is trading volume and SE denotes stock exchange.

Again, the 10-fold cross-validation was used to obtain the average merits (Table 9). The log of market capitalization and type of stock exchange were the most important predictors of stock return volatility. However, the length of news articles and positive sentiment also appear to be relevant indicators. On the contrary, the merits of negative and overall sentiment were negative, suggesting that their contributions to prediction models are small. The merits of the remaining technical indicators were positive.

Conclusion

Prior literature reported an important role of text analysis in predicting stock market indicators such as stock prices, trading volumes and volatility. The present study was designed to develop a novel meta-learning model for predicting abnormal stock return volatility. Both the financial indicators of stock market and the textual analysis of news articles were included in this model. The meta-learning approach was used to model the decision-making process of various investors. The results of the prediction model suggest that this approach significantly improves

the prediction accuracy of abnormal stock return volatility. Returning to the hypothesis posed at the beginning of this study, it is now possible to state that it is more effective to combine the prediction capacities of multiple classifiers to obtain a more accurate model. A high complexity of the investigated problem was documented by the necessity to use rotated feature spaces in Rotation trees for each base predictors. The study has also gone some way towards enhancing our understanding of predicting stock return volatility. An important finding is that it seems to be the length of news articles that has the most important impact on future stock return volatility. This was also confirmed by its high merit in prediction models.

Our findings may have broader important implications for related issues such as the prediction of stock prices and trading volumes. Meta-learning approaches combining technical indicators and text analysis may improve the prediction accuracy in these financial problems. However, a number of caveats need to be noted regarding the present study. The most important limitation lies in the fact that a relatively short period of time and only the largest U.S. firms were examined. Therefore, caution must be applied, as the findings might not be transferable to small and medium firms and different macroeconomic context. Further experimental investigations are needed to obtain more generalizable findings. It is also strongly recommended to investigate the role of market psychology (sentiment) indicators in future studies. These evaluate the general opinions of investors' population on future stock market development. For example, indicators such as Short Sale, Most Active Stock Indicator or confident indices can be applied to reflect investment optimism not only on stock markets but also on bond markets because bond market sentiment usually affect future stock market development. This additional information may further improve the performance of the proposed prediction model.

Acknowledgements

This work was supported by the scientific research project of the Czech Sciences Foundation Grant No: 16-19590S. The assistance on data collection from K. Sindelkova is gratefully acknowledged.

References

- [1] Ahern, K. R., Sosyura, D. (2014). Who writes the news? Corporate press releases during merger negotiations. *The Journal of Finance* 69(1), 241–291.
- [2] Antweiler W., Frank M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* 59(3), 1259–1294.
- [3] Ballings, M., Van den Poel, D., Hespels, N., Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications* 42(20), 7046–7056.
- [4] Barber, B., Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785–818.
- [5] Breiman L. (1994) *Bagging predictors*. Technical Report no. 421, Department of Statistics, University of California, Berkeley.
- [6] Breiman, L. (1996). Bagging predictors. *Machine Learning* 24(2), 123–140.
- [7] Bryll, R. (2003). Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition* 20(6), 1291–1302.
- [8] Clatworthy, M., Jones, M. J. (2003). Financial reporting of good news and bad news: evidence from accounting narratives. *Accounting and Business Research* 33(3), 171–185.
- [9] Da, Z., Engelberg, J., Gao, P. (2011). In search of attention. *Journal of Finance* 66(5), 1461–1499.

- [10] Davis, A. K., Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&A. *Contemporary Accounting Research* 29(3), 804–837.
- [11] Davis, A. K., Piger, J. M., Sedor, L. M. (2012). Beyond the numbers: measuring the information content of earnings press release language. *Contemporary Accounting Research* 29(3), 845–868.
- [12] DellaVigna, S., Pollet, J. (2009). Investor inattention and Friday earnings announcements, *Journal of Finance* 64(2), 709–749.
- [13] Demers, E., Vega, C. (2010). *Soft information in earnings announcements: News or noise?* INSEAD Faculty & Research Working Paper 2010/33/AC, available at <http://www.insead.edu/facultyresearch/faculty/personal/edemers/documents/SoftInformationinEarningsAnnouncementsNewsorNoise-INSEADWP.pdf>
- [14] Doran, J.S., Peterson, D.R., Price, S.M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics* 45(2), 402–434.
- [15] Drasnar, G. (1995). *Hazardní hry: úvod do spekulace s cennými papíry*. Prague: Academica, in Czech.
- [16] Edwards, R. D., Magee, J., & Bassetti, W. C. (2007). *Technical analysis of stock trends*. CRC Press.
- [17] Fama E.F., French K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.
- [18] Fang, L. H., Peress, J. (2009). Media coverage and the cross-section of stock returns, *The Journal of Finance* 64(5), 2023–2052.
- [19] Freund, Y., Schapire, R. E. (1996). Experiments with a new boosting algorithm. In: *Proc. of the 13th Int. Conf. on Machine Learning*, San Francisco, pp.148–156.
- [20] Freund Y., Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1), 119–139.
- [21] Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance* 68(3), 1267–1300.
- [22] Grimmer, J., Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3), 267–297.
- [23] Hajek, P., Bohacova, J. (2016). Predicting abnormal bank stock returns using textual Analysis of Annual Reports - A neural network approach. In: Jayne, C., Iliadis, L. (eds) *Engineering Applications of Neural Networks*. EANN 2016. Communications in Computer and Information Science 629. Springer, Cham, pp. 67–78.
- [24] Hajek, P., Olej, V. (2013). Evaluating sentiment in annual reports for financial distress Prediction Using Neural Networks and Support Vector Machines. In: Iliadis, L., Papadopoulos, H., Jayne, C. (eds.), *Engineering Applications of Neural Networks*. EANN 2013. Communications in Computer and Information Science 384. Springer, Berlin, Heidelberg, pp. 1–10.
- [25] Hajek, P., Olej, V., Myskova, R. (2015). Predicting financial distress of banks using random subspace ensembles of support vector machines. In: Silhavy, R., Senkerik, R., Oplatkova, Z., Prokopova, Z., Silhavy, P. (eds.), *Artificial Intelligence Perspectives and Applications*. Advances in Intelligent Systems and Computing 347. Springer, Cham, pp. 131–140.
- [26] Hajek, P., Olej, V., Myskova, R. (2014). Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy* 20(4), 721–738.

- [27] Henry, E. (2008). Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication* 45(4), 363–407.
- [28] Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8), 832–844.
- [29] Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *Accounting Review* 89(3), 1083–1113.
- [30] Li, F. (2010). The information content of forward-looking statements in corporate filings - a naïve machine learning approach. *Journal of Accounting Research* 48(5), 1049–1102.
- [31] Kira, K., Rendell, L.A. (1992). A practical approach to feature selection. In: *Proc. of the 9th International Workshop on Machine Learning*, pp.249–256.
- [32] Kothari, S. P., Li X., Short J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review* 84(5),1639–1670.
- [33] Loughran T., Mcdonald B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66(1), 35–65.
- [34] Meluzin, T., Zinecker, M. (2014). Reasons for IPO Implementation: Empirical Evidence from the Polish Capital Market. *Engineering Economics*, 25(3), 294-301.
- [35] Meluzin, T., Zinecker, M. (2016). Trends in IPOs: The Evidence from CEE Capital Markets. *Equilibrium* 11(2), 327-341.
- [36] Melville P., Mooney, R. (2003). Constructing diverse classifier ensembles using artificial training examples. In: *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence*, Acapulco, Mexico, pp. 505–510.
- [37] Myskova, R., Hajek, P. (2016). The effect of managerial sentiment on market-to-book ratio. *Transformation in Business & Economics* 15(2A), 498–513.
- [38] Powers, D. M. W. (2011). Evaluation from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 1(2), 37–63.
- [39] Price, S., Doran, J., Peterson, D., Bliss, B. (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. *Journal of Banking and Finance* 36(4), 992–1011.
- [40] Rodriguez, J. J., Kuncheva, L. I., Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(10), 1619–1630.
- [41] Tetlock P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139–1168.
- [42] Tsai, C. F., Lin, Y. C., Yen, D. C., Chen, Y. M. (2011). Predicting stock returns by classifier ensembles. *Applied Soft Computing* 11(2), 2452–2459
- [43] Webb, G. I. (2000). MultiBoosting: A technique for combining boosting and wagging. *Machine Learning* 40(2), 159–196.
- [44] Witten I. H., Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd ed.
- [45] Zhou, S., Lai, K. K., Yen, J. (2012). A dynamic meta-learning rate-based model for gold market forecasting. *Expert Systems with Applications* 39(6), 6168–6173.