TRENDY EKONOMIKY A MANAGEMENTU
TRENDS ECONOMICS AND MANAGEMENT
ISSN 1802-8527 (Print) / ISSN 2336-6508 (Online)

2024 42(1): 81–94
DOI: http://dx.doi.org/10.13164/trends.2024.42.81

Detection of Fake News Using a Machine Learning Approach

Zuzana Janková, Michal Páleš, Marcin Potrykus, László Buics

Abstract

Purpose of the article: The tremendous rise of artificial intelligence and the use of ChatGPT enables the unprecedented rapid spread of various messages. The fundamental issue at present consists in the spread of false, unverified news and information known as fake news. This great challenge of the modern world causes deliberate manipulation of public opinion, can contribute to the loss of value in the stock market and poses many risks on a global level.
Methodology/methods: For the detection of fake news, a dataset containing text messages from the social platform X is used. Within the framework of natural language processing (NLP), a text analysis was performed including traditional pre-processing steps, such as tokenization, removal of traces of words, excessive punctuation, *etc.* Subsequently the Bag of Words transformation was used, the text is coded into word vectors through the word embedding method and Word2Vec vectorization. Based on previous research and impressive practical performance, the Support Vector Machine (SVM) technique is chosen for fake news classification, which is a highly robust and effective machine learning algorithm.
Scientific aim: The issue of detection and classification of information disseminated on online platforms is difficult, and so far, no unambiguous approach has been provided that would provide satisfactory results. To fill the research gaps, this paper is focused on the detection of fake news from messages published on a social network using Machine Learning (ML) methods, specifically, the SVM algorithm is chosen.
Findings: Preprocessing the text data reduced the dimensionality of the dataset by almost 50%, as many news headlines contained a large number of meaningless tokens or excessive punctuation. The importance of this step has been proven, especially when using unstructured data from social platforms. The accuracy of fake news classification is almost 75% using SVM. The contribution lies in expanding the current literature on approaches to fake news detection, particularly through tuning the SVM model by searching for hyperparameters that minimize cross-validation loss.
Conclusions: Due to the fact that unstructured data and individual pre-processing steps can distort certain elements revealing fake news from legitimate news, it is advisable to focus in more detail on individual pre-processing steps in future research. In particular, excessive punctuation or frequent use of stop words can provide additional elements that can help separate fake news from real news.

Keywords: fake news, false information, financial market, stock market, social media, SVM, ML

JEL Classification: C45, G11, G12

Introduction

The rapid growth of online platforms and community activity has accelerated the spread of unverified, false information and hoaxes referred to as "fake news", which not only plagues society in general, but can have serious political but also economic consequences and social impacts on the public and private sectors. Especially in the financial market, combating fake news in the era of social media is a major issue and represents a serious threat to investors. Above all, social media such as X or Meta are criticized for promoting articles that later turned out to be false. One can come across many definitions of fake news. McGonale (2017) refers to fake news as news articles that are intended to deceive or misinform others. According to Allcott and Gentzkow (2017), this is intentionally and verifiably false information that is created to mislead the reader. According to Egelhofer and Lecheler (2019), news with low authenticity published by conventional media is also labelled as fake news.

According to Clarke *et al.* (2021), the reach of such reports is astounding, as Vosoughi *et al.* (2018), fake news is 70% more likely to be retweeted than legitimate news. Moreover, the real impact on the financial market is not fully understood. In practice, one can come across promotional companies that deliberately create bullish articles, or conversely news that deliberately try to damage the good name or brand of a competing company. Hong *et al.* (2023a) draw attention to the fundamental importance of verifying the validity and authenticity of published information, especially in times of crisis. In these situations, rumours and fake news spread very easily and the public does not realize or verify the truth of the information. Uncertainty caused by misinformation is essentially ex post uncertainty and can have significant economic consequences. All these motivations forced the European Commission to promulgate a "Code of Practice on Disinformation". According to Lee *et al.* (2024), the Code's guidelines are increasingly adopted by operators of online social platforms. With the progress of artificial intelligence (AI), another threat arises, namely, the increasing activity of social bots on social platforms (Hajli *et al.*, 2022). These are autonomous actors controlled by algorithms and software that publish content on these platforms. It is striking that more than a third of tweets come from social bots. Recent research by Shi *et al.* (2019) shows that this malicious type of bot spreads false information and misinformation and manipulates the stock market.

According to the principles of the efficient market hypothesis, fake news should not affect market prices, as this false information should be rejected by rationally thinking subjects in an efficient market (Fama, 1970). This means that the market should not react to false information, as its impact on the markets is irrational (Fong, 2021). Despite the significant discrepancy between empirical evidence and economic theory, there is no relevant model providing an explanation for the persistence of fake news in financial markets. This great challenge of the modern world causes deliberate manipulation of public opinion, can contribute to the loss of value in the stock market, and poses many risks on a global level. The process of detecting fake news is time- and labour-intensive and represents a major challenge in the current world of digital transformation (Lee *et al.*, 2024). Despite the growing importance of detecting fake news published mainly on social networks, this area has received a minority of academic research attention, probably due to lack of data and identification problems, as added by Wang and Ye (2023), especially in the context of financial markets, which are an inherent part of the economic system of countries, this issue is insufficiently covered.

To fill the research gaps, this paper is focused on detecting fake news from messages published on a social network using Machine

Learning (ML) methods. The growing progress of Industry 4.0, the development of artificial intelligence could bring tools that can automatically identify and reveal such information using linguistic characteristics.

1. Literature review

At the beginning, it is appropriate to state that information can be characterized using two elements "strength" and "weight" according to Antoniou *et al.* (2017). The first element represents the tone of the message and the extremity of the information. The second element represents the quality of information. Investors weight these two elements differently, mainly depending on the external environment. According to Wu *et al.* (2022) the stronger the intensity of monitoring of news publishing, the more resistant investors are to false information. In other words, they rely more on "weight" than "strength" information. Hong *et al.* (2023a) confirms this idea that publicity surveillance significantly reduces the impact of misinformation. On the contrary, in the case of a favourable situation in the development of financial markets, investors are less vigilant and look more at the "strength" of information.

As it was already mentioned in the introductory part, the fact that the amount of available research on the impact of false information on financial markets is limited also testifies to the insufficient research of the issue in question. According to Fong (2021), the vast majority of research published so far focuses on the political and social consequences of the spread of disinformation, which are relatively separated from the fight against this information and the methods of its detection. According to a search in the Web of Science after entering the sequence of keywords *("false new*" OR "fake new*" OR "disinformation" OR "misinformation" OR "false information") AND ("stock market" OR "financial market")* only 35 articles were generated and through the PRISMA methodology shown in Figure 1, the relevant articles dropped to 20. It is evident that, although the literature on the effect of fake news on financial market movements remains limited, it has yielded some interesting insights. Wang and Ye (2023) investigated the effect of fake news extracted from social networks on post-earnings announcement drift (PEAD). PEAD reflects how stock prices and investors interpret and react to unexpected earnings news. According to the authors' output, companies can play an active role in improving information efficiency and in the fight against fake news on the stock market. Clarifications of fake news by companies help to improve the information environment. Similar results can be found in a study by Xu (2021), who reports that if firms immediately respond and counter fake news, they reduce the probability of future attacks by 19% and significantly reduce the negative cumulative impact on stock returns. Clarke *et al.* (2021) confirm that fake news attracts investors' attention more than legitimate news. Moreover, they found that neither the commentators nor the editors are able to detect such messages. The authors focused on the website Seeking Alpha, which is one of the largest social platforms for financial markets. Drainage learning algorithms represent a powerful tool to successfully identify fake news. Hong *et al.* (2023a) focused on the period of the pandemic situation of COVID-19 and investigated the relationship between fake news and the occurrence of extreme risks in the stock markets of selected developed and developing countries. According to the authors' output, interest in fake news is greater in developed countries compared to developing countries. After the thunder of legitimate information, the volatility of stock markets decreases. Clarke *et al.* (2021) observed abnormal trading volume at the time of fake news publication. Hong *et al.* (2023a) and Clarke *et al.* (2021) agreed that when legitimate information

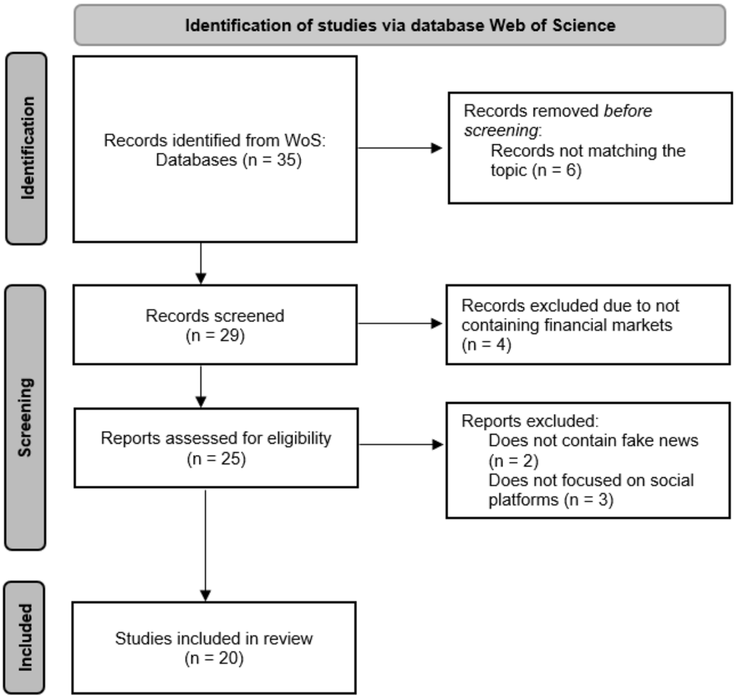

Figure. 1. PRISMA diagram of literature review. Source: own research.

is included, the volatility of stock markets decreases.

In their research, Chen and Chen (2024) focused on the potential impact of misinformation published online on stock returns. Li *et al.* (2018) found that institutional investors can profit from fake news detection. Fong (2021) confirms that prolonged fake news can suppress appropriate investor responses to subsequent legitimate information. Lee *et al.* (2024) use Long Short-Term Memory (LSTM) deep learning techniques to improve the accuracy of misinformation detection and further investigate the relationship between the emotions evoked by messages and the veracity of such messages. Moreover, the authors demonstrate that news from the US media has an impact on stock markets with important implications for investors from emerging economies. Chung *et al.* (2023) apply a novel deep learning approach (TRNN) to the detection of false information. TRNN is deep learning with augmentation based on psychological and social theories to detect misinformation. Hong *et al.* (2023b) focused on the Chinese stock market and investigated the influence of "black mouth", a form of manipulation through misinformation. According to the authors, this phenomenon leads to abnormal investor attention and triggers abnormal stock returns. Kogan *et al.* (2023) state that in a situation where the cost of information is high and the ability to take corrective action is limited, the setup of financial markets can encourage the wider impact of misinformation. Additionally, they found that social media messages are associated with retail price fluctuations and trading activity, as retail investors are the primary participants in these social platforms.

2. Methodology

Fake news detection may be divided into two broad categories, *i.e.* (i) manual (ii) automated (Taher *et al.*, 2022). The first category is based on comparing unverified reports with truly legitimate information. The second category uses automated systems to detect false information, such as a combination of natural language processing (NLP) and artificial intelligence (AI), which can include machine learning (ML) and deep learning (DL) algorithms.

Natural Language Processing (NLP)

1) Text pre-processing
   Pre-processing unstructured text data is one of the most important steps in NLP. Text messages posted on social networks consist of a variety of content, are raw and noisy. For that reason, the text needs to be thoroughly cleaned and pre-processed, as it contains meaningless information such as various tags or links (Özöğür Akyüz *et al.*, 2024). As part of this first step of natural language processing, the text is tokenized, then the text is converted to lowercase letters, stop words such as conjunctions, prepositions, articles, *etc.*, but also numbers and punctuation marks, last but not least, web links and special characters are removed. which are difficult to understand and often contain ignored words that lack meaning and unnecessarily increase the size of the text (Janková, 2023).
2) Bag of Words
   Bag of Words (BoW) is an approach to the representation of documents whose tokens are characterized as tokens that are assigned a weight corresponding to the value of a character, or presence of words in the document through the term frequency approach. The higher the weight assigned, the more important that token is compared to a token with a lower weight (Khan *et al.*, 2020).
3) Text encoding
   Text pre-processing is followed by encoding, which generates a set of vector elements that represent individual words in an unstructured data set.
4) Word2Vec
   Word2Vec is a more advanced form of word representation, providing word embeddings that take into account aspects such as linguistic similarities – and thus provide better classification results, see (Fazlija, Harder, 2022). Basically, it is a two-layer neural network for processing natural language and representing words as a set of vectors. After preprocessing, the text is fed into Word2vec to encode and transform the text into word vectors (Liu *et al.*, 2022).

Support Vector Machine (SVM)

Machine Learning (ML) and Artificial Neural Networks (ANN) methods are widely used to classify unstructured data. The review article by Janková (2023) shows the most common use of methods for classifying text data in the context of stock markets. Her article shows that more than 75% of the research analysed applies the Support Vector Machine (SVM) method to classification. Within ML techniques, it is necessary to (i) split the data set into a training and a test set; (ii) select a classifier model and train it on the training dataset; (iii) classifying the new text using a well-trained classifier on the test dataset.

Specifically, for the purposes of this paper, the SVM technique, which is a highly robust and efficient machine learning algorithm, is chosen, even according to previous research and impressive practical performance. SVM has demonstrated excellent performance across a range of domains including financial markets. SVM can be used for both classification and regression tasks. According to Dahal *et al.* (2024) SVM classifies each x_i do class input in the range 1 to p by identifying the hyperplane that separates the different data types. A hyperplane represents a flat hyperspace of dimension $p – 1$ in the case of a p-dimensional space. If we consider an input matrix X, which is structured as an nxp matrix, where n indicates the number of observations in the data set, p indicates the number of features. The resulting vector is labeled as Y and falls into one of the classes (–1, 1). The hyperlane by which the SVM predicts the class is described by the equation:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = 0 \qquad (1)$$

where

$$\beta = (\beta_0, \beta_2, \ldots, \beta_p)^T$$

represents a vector of parameters that can be estimated through an optimization method with the following property:

(i) if $y = 1$, then

$$(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p > 0;$$

(ii) if $y = -1$, then

$$(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p < 0.$$

The SVM technique offers different types of kernels (linear, polynomial, and radial) and hyperparameters (gamma and cost), which work together to improve the performance of this technique and are the builders of accurate output. According to Dahal *et al.* (2024), the tuning of these parameters is essential for the performance of SVM, as incorrect setting and balancing of parameters can cause overfitting or underfitting (Wang *et al.*, 2018).

Classifier Evaluation

Evaluation of classification accuracy also represents a necessary element of the classification process. The confusion matrix is the most commonly used tool to measure the performance of machine learning classifiers. First, it is necessary to present the individual metrics that are determined from the confusion matrix (Sangsavate *et al.*, 2019).

- True positive (*TP*) positive tuples that are correctly classified.
- True negative (*TN*) refers to negative tuples that are correctly assigned to a class.
- False positive (*FP*) represents positive tuples that are not correctly labelled through the assigned class.
- False negative (*FN*) is a negative tuple that is not correctly labelled via a classification category.

Table 1 shows the confusion matrix. The individual columns represent the prediction classes, while the rows correspond to the current class.

Various metrics can be calculated from the confusion matrix. For the purposes of this paper, accuracy is most important. The count is defined as the sum of the number of true positive (*TP*) and true negative (*TN*) divided by the sum of the number of true negative (*TN*), true positive (*TP*), false negative (*FN*), and false positive (*FP*):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (2)$$

3. Proposed model

The processing procedure is shown in the diagram (Figure 2). First, the data is obtained, followed by their pre-processing according to NLP procedures and the division of the data set into training and test sets. On the training set, the SVM model is trained, tuned and then verified on the test data set. Finally, a confusion matrix evaluating the accuracy of the model for classifying fake news is compiled.

Table. 1. Confusion matrix.

Predication/ Actual class	**Negative**	**Positive**
False	False Negative (*FN*)	False Positive (*FP*)
True	True Negative (*TN*)	True Positive (*TP*)

Source: own research.

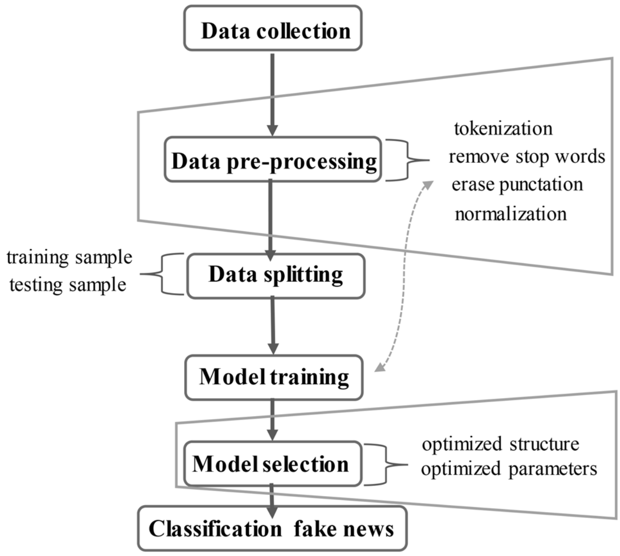

Figure. 2. Workflow of classification fake news with supervised learning. Source: own research.

Textual data

The unstructured text dataset is obtained from the Kaggle database. This is the Golovin (2022) fake news dataset, which comes from the FakeNewsNet portal. The text data includes the title of the article, the url of the article, the source web domain on which the article was posted, the number of retweets and the class designation: 1 indicates legitimate news, 0 indicates fake news. A sample of eight articles is shown in Figure 3. The unstructured data is divided into two data sets in the ratio of 70% training set, 30% test set. In total, the dataset contains almost 22,000 unique messages. Figure 4 shows the number of distribution of headlines in the entire data set within the breakdown into legitimate and fake news. It can be seen from the figure that a large part, almost 18,000 messages from the dataset contain legitimate information, while less than 6,000 are detected as fake news.

Text processing

All calculations are performed in MATLAB R2021a software. Text preprocessing is done in several steps. First, tokenization is performed as a basic building block of NLP through the `tokenizedDocument` command. Then the traditional and absolutely necessary steps are taken to clean up the text and reduce its dimensionality, such as removing stop words `removeStopWords`, removing excessive punctuation `erasePunctuation`,

untitled.mlx | ans = 8×5 table

	title	news_url	source_domain	tweet_num	real
1	'Kandi Burruss Explodes Over Rape Accusation on 'Real Housewives of Atlanta' Reunion (Video)'	'http://toofab.c…	'toofab.com'	42	1
2	'People's Choice Awards 2018: The best red carpet looks'	'https://www.to…	'www.today.com'	0	1
3	'Sophia Bush Sends Sweet Birthday Message to 'One Tree Hill' Co-Star Hilarie Burton: 'Breyton 4eva''	'https://www.et…	'www.etonline.com'	63	1
4	'Colombian singer Maluma sparks rumours of inappropriate relationship with AUNT'	'https://www.d…	'www.dailymail.co.uk'	20	1
5	'Gossip Girl 10 Years Later: How Upper East Siders Shocked the World and Changed Pop Culture F…	'https://www.z…	'www.zerchoo.com'	38	1
6	'Gwen Stefani Got Dumped by Blake Shelton Over "Jealousy and Drama" (EXCLUSIVE)'	'www.intouchw…	'www.intouchweekly.c…	45	0
7	'Broward County Sheriff Fired For Lying About Parkland'	'https://yourne…	'yournewswire.com'	124	0
8	'Amber Rose Shuts Down French Montana Dating Rumors, Calls Rapper Her 'Bruvaaa''	'www.etonline.…	'www.etonline.com'	4	0

Figure. 3. A sample text dataset. Source: own research.

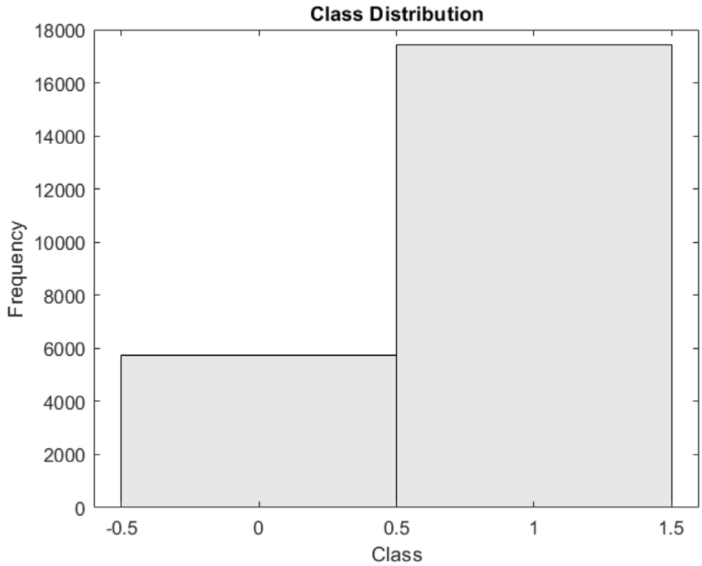

Figure. 4. Distribution of fake news and legitimate news in the dataset. Source: own research.

normalizing `normalizeWords`, or morphological markers can be added to tokens `addPartOfSpeechDetails`.

The next step is to shorten the documents. For best results, the target length should be short without removing large amounts of data. Therefore, the histogram (Figure 5) of the lengths of the training document is shown. Most training titles have around 10 to 15 tokens.

After preprocessing, the number of words was counted, or modified text and raw text tokens. By dividing them and subtracting them from one, the reduction percentage was

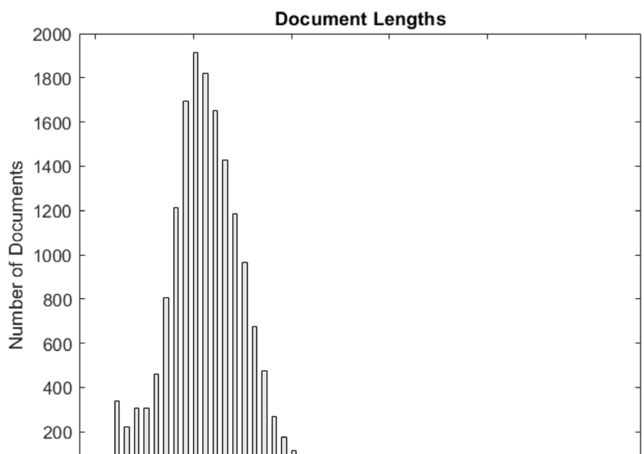

Figure. 5. Document length histogram of the training dataset. Source: own research.

created. As part of the pre-processing, the text was reduced by almost 0.4940, in other words, almost half of the text was reduced, as it mainly contained a large number of stop words and punctuation.

After the text has been pre-processed, it is subsequently transformed into a numerical form using `bagOfWords`, and then the text is encoded by `encode`. Through the Text Analytics Toolbox and the fastText English, 16 Billion Token Word Embedding support package installed. The word embedding package is one of the popular representations of vocabulary, in the case of the MATLAB package it contains 16 trillion tokens. Owing to this package, it is possible to capture the context of the words in the document, its syntactic and semantic similarity, including its relationship to other words in the document. The package installation verification can be performed with `emb = fastTextWordEmbedding`. Using the word to embedding vector `word2vec` mapping technique, a shallow neural network is trained to embed words.

Support Vector Machine

Due to the fact that the data set contains two classes – legitimate news and fake news, the `fitcsvm` command can be used for classification, which is applied for binary classification through the support vector machine. This command supports data mapping through kernel functions. For the thinning process, the SVM classifier is set to gaussian basis kernel "KernelFunction","gaussian", the software is left to find the scale value for the kernel function "Standardize", true, and the predictors are standardized to "KernelScale","auto".

Figure 6 and Figure 7 contain the outputs of setting the SVM classifier. Through automatic optimization of hyperparameters, hyperparameters minimizing cross-validation loss are sought.

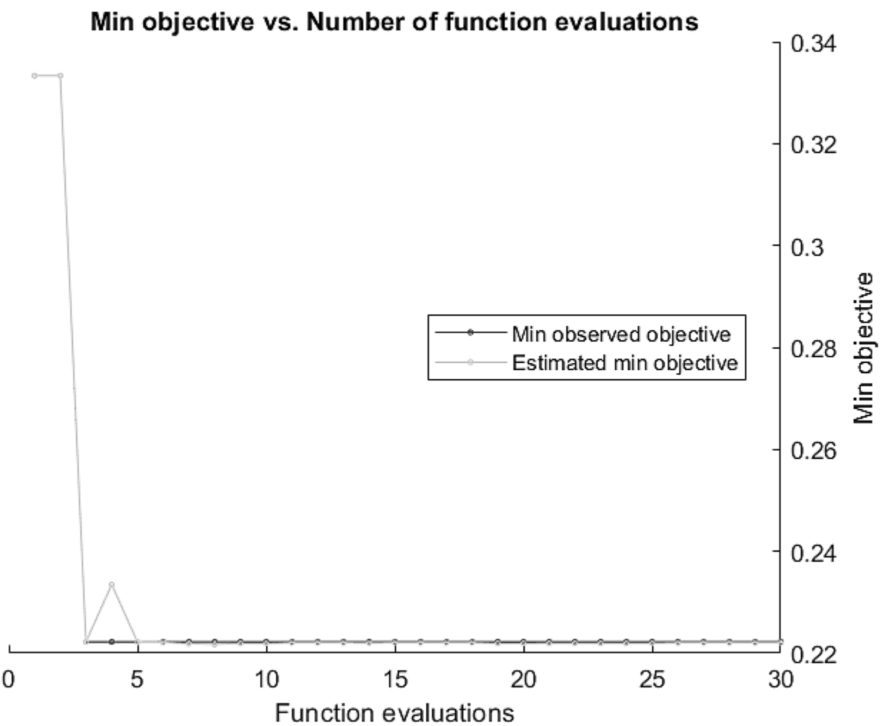

Figure. 6. SVM function evaluations. Source: own research.

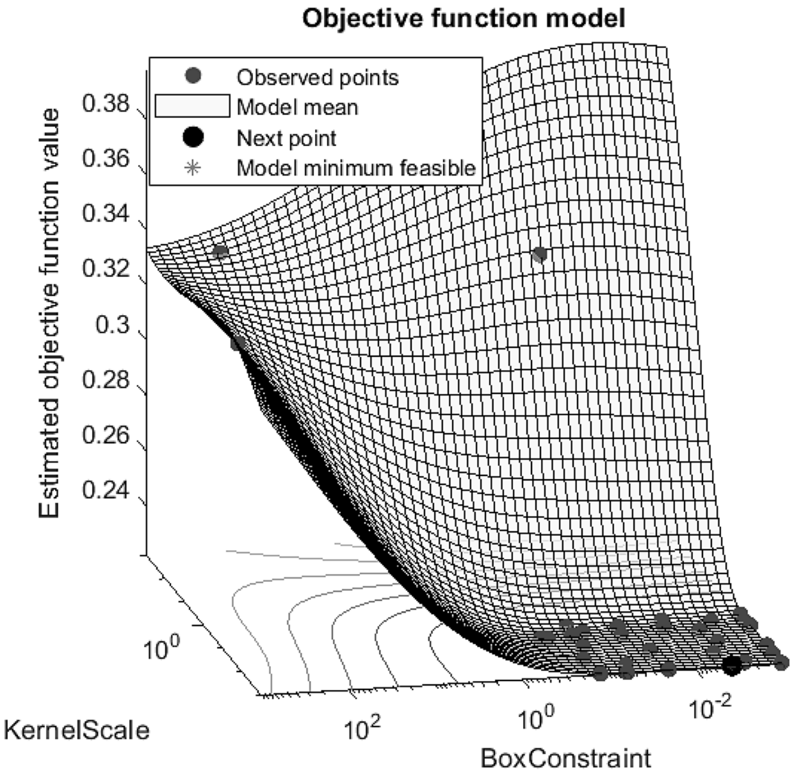

Figure. 7. Objective function SVM model. Source: own research.

The accuracy of the SVM model is calculated using the confusion matrix. Accuracy is 74.9928%, which is a relatively solid accuracy of fake news classification.

4. Conclusion

The paper dealt with the detection and classification of messages marked as fake news. The dataset of messages posted on the social network X was obtained from the Kaggle database. Text unstructured data was pre-processed using natural language processing according to traditional methods. Furthermore, the technique of inserting words and converting individual tokens into vector space was applied. Through the machine learning algorithm, namely the Support Vector Machine, which was trained and tuned, and the subsequent accuracy and performance of the classifier was verified on the test data set. The accuracy of the classifier was calculated from the confusion matrix, which is 74.99%. The accuracy of the classification, due to the fact that it is unstructured data, and the recognition of fake news may not be completely clear from the context of the message, almost 75% accuracy can be considered a solid result. In comparison to other studies such as Horne *et al.* (2017) used SVM for multi-class classification, and in addition to identifying fake news, they also used a satire scale. Their classifier achieved an accuracy of 78%. Similarly, Sivasankari and Vadivu (2022) focused on detecting fake news on Twitter. They used SVM to detect with an accuracy of 84.2%. Burgoon *et al.* (2003) used C4.5 Decision Tree with 15-fold cross--validation and created a hierarchical tree structure. The accuracy of their model was 60.72%. Hakak *et al.* (2021) applied Random Forest (RF) to detect fake tweets that were published during the US presidential election, the accuracy of the classifier was 85%. However, in recent years, deep models (DL) have come to the fore. Nassif *et al.* (2022) use convolutional neural networks to classify news into authentic and fraudulent, their outputs show an accuracy of over 96%. Al-Asadi and Tasdemir (2022) went a step further and combined SVM, CNN and LSTM, their model fusion shows excellent results with an accuracy of 98%. Thus, in further research, a similar model fusion can be strongly considered, as in the case of Al-Asadi and Tasdemir (2022). It would also be appropriate in subsequent research to focus on the individual steps of preprocessing, for example, whether the removal of excessive punctuation is not a crucial element for recognizing fake news and omitting this step would increase the accuracy of detecting and classifying fake news.

Acknowledgment

The paper was compiled with the support of the project VEGA 1/0431/22 entitled "Implementácia inovatívnych prístupov modelovania rizík v procese ich riadenia v interných modeloch poisťovní v kontexte s požiadavkami direktívy Solvency II" provided by the Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

Al-Asadi, M. A., Tasdemir, S. (2022). Using Artificial Intelligence Against the Phenomenon of Fake News. *A Systematic Literature Review*, 1001(1), pp. 39–54. DOI:10.1007/978-3-030-90087-8_2.

Allcott, H., Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), pp. 211–36. DOI:10.1257/jep.31.2.211.

Antoniou, C., Harrison, G. W., Lau, M. I., Read, D. (2017). Information Characteristics and Errors in Expectations: Experimental Evidence. *Journal of financial and quantitative analysis*, 52(2), pp. 737–750. DOI:10.1017/S0022109017000035.

Burgoon, J. K., Blair, J. P., Qin, T., Nunamaker, J. F., Zeng, D., Miranda, R., Schroeder, J., Demchak, C., Chen, H., Madhusudan, T. (2003). Detecting deception through linguistic analysis. *In INTELLIGENCE AND SECURITY INFORMATICS, PROCEEDINGS*, 2665, pp. 91–101. Springer Nature. DOI :10.1007/3-540-44853-5_7.

Clarke, J., Chen, H., Du, D., Hu, Y. J. (2021). Fake News, Investor Attention, and Market Reaction. *Information Systems Research*, 32(1), pp. 35–52. DOI:10.1287/isre.2019.0910.

Chen, T. H., Chen, K. S. (2024). The effect of investor attention on stock price crash risk. *Journal of Empirical Finance*, 75, pp. 101456. DOI:10.1016/j.jempfin.2023.101456.

Chung, W., Zhang, Y., Pan, J. (2023). A Theory-based Deep-Learning Approach to Detecting Disinformation in Financial Social Media. *Information systems frontiers,* 25, pp. 473–492. DOI :10.1007/s10796-022-10327-9.

Dahal, K. R., Gupta, A., Pokhrel, N. R. (2024). Predicting the Direction of NEPSE Index Movement with News Headlines Using Machine Learning. *Econometrics*, 12(2), p. 16. DOI:10.3390/econometrics12020016.

Egelhofer, J. L., Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, 43(2), pp. 97–116. DOI:10.1080/23808985.2019.1602782.

Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance, 25, pp. 383–417.

Fazlija, B., Harder, P. (2022). Using Financial News Sentiment for Stock Price Direction Prediction. *Mathematics*, 10(13), pp. 2156. DOI:10.3390/math10132156.

Fong, B. (2021). Analysing the behavioural finance impact of fake news' phenomena on financial markets: A representative agent model and empirical validation. *Financial Innovation*, 7(1), pp. 1–30. DOI:10.1186/s40854-021-00271-z.

Golovin, A. (2022). Fake News. Kaggle. DOI:10.34740/kaggle/ds/2623949. Retrieved from: https://www.kaggle.com/ds/2623949.

Hajli, N., Saeed, U., Tajvidi, M., Shirazi, F. (2022). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. *British journal of management*, 33(3), pp. 1238–1253. DOI:10.1111/1467-8551.12554.

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future generation computer systems*, 117, pp. 47–58. DOI:10.1016/j.future.2020.11.022.

Hong, Y., Qu, B., Yang, Z., Jiang, Y. (2023a). The contagion of fake news concern and extreme stock market risks during the COVID-19 period. *Finance research letters*, 58, pp. 104258. DOI:10.1016/j.frl.2023.104258.

Hong, Z., Liu, Q., Tse, Y., Wang, Z. (2023b). Black mouth, investor attention, and stock return. *International review of financial analysis*, 90, pp. 102921. DOI:10.1016/j.irfa.2023.102921.

Horne, B. D., Adali, S. (2017). This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. DOI:10.48550/arxiv.1703.09398.

Janková, Z. (2023). CRITICAL REVIEW OF TEXT MINING AND SENTIMENT ANALYSIS FOR STOCK MARKET PREDICTION. Journal of Business Economics and Management, 24(1), pp. 177–198. DOI:10.3846/jbem.2023.18805.

Khan, W., Malik, U., Ghazanfar, M. A., Azam, M. A., Alyoubi, K. H., Alfakeeh, A. S. (2020). Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft computing (Berlin, Germany)*, 24(15), pp. 11019–11043. DOI:10.1007/s00500-019-04347-y.

Kogan, S., Moskowitz, T. J., Niessner, M. (2023). Social Media and Financial News Manipulation. *Review of Finance*, 27(4), pp. 1229–1268. DOI:10.1093/rof/rfac058.

Lee, C. -Chiang, Chen, T. -Hsuan, Yen, M. -feng. (2024). The Dissemination of Politically Motivated False Information and Its Influence on the Stock Market. *Emerging markets finance & trade*, 60(9), pp. 2048–2067. DOI:10.1080/1540496X.2023.2298266.

Li, Q., Wang, J., Bao, L. (2018). Do institutions trade ahead of false news? Evidence from an emerging market. *Journal of Financial Stability*, 36, pp. 98–113. DOI:10.1016/j.jfs.2018.02.001.

Liu, Q., Wang, X., Du, Y. (2022). The weekly cycle of investor sentiment and the holiday effect – An empirical study of Chinese stock market based on natural language processing. *Heliyon*, 8(12), e12646-e12646. DOI:10.1016/j.heliyon.2022.e12646.

McGonagle, T. (2017). "Fake news" false fears or real concerns? *Netherlands Quarterly of Human Rights*, 35(4), pp. 203–209. DOI:10.1177/0924051917738685.

Nassif, A. B., Elnagar, A., Elgendy, O., Afadar, Y. (2022). Arabic fake news detection based on deep contextualized embedding models. *Neural computing & applications*, 34(18), pp. 16019–16032. DOI:10.1007/s00521-022-07206-4.

Özöğür Akyüz, S., Karadayı Ataş, P., Benkhaldoun, A. (2024). Predicting stock market by sentiment analysis and deep learning. *Operations research and decisions*, 34(2). DOI:10.37190/ord240206.

Sangsavate, S., Tanthanongsakkun, S., Sinthupinyo, S. (2019). Stock Market Sentiment Classification from FinTech News. Online. *In: 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE). IEEE*, 2019, pp. 1–4. DOI:10.1109/ICTKE47035.2019.8966841.

Shi, P., Zhang, Z., Choo, K. -K. R. (2019). Detecting Malicious Social Bots Based on Clickstream Sequences. *IEEE access*, 7, pp. 28855–28862. DOI:10.1109/ACCESS.2019.2901864.

Sivasankari, S., Vadivu, G. (2022). Tracing the fake news propagation path using social network analysis. *Soft computing (Berlin, Germany)*, 26(23), pp. 12883–12891. DOI:10.1007/s00500-021-06043-2.

Taher, Y., Moussaoui, A., Moussaoui, F. (2022). Automatic Fake News Detection based on Deep Learning, Fast Text and News Title. *International journal of advanced computer science & applications*, 13(1), pp. 146–158. DOI:10.14569/IJACSA.2022.0130118.

Vosoughi, S., Roy, D., Aral S. (2018). The Spread of True and False News Online. *Science*, 359, pp. 1145–1151.

Wang, M., Ye, T. (2023). Do Firms' Social Media Fake News Clarifications Mitigate the PEAD Anomaly? Evidence from a Policy Experiment in China. *Emerging markets finance & trade*, 59(5), pp. 1281–1299. DOI:10.1080/1540496X.2022.2128755.

Wang, S., Qiang, L., En Z-, Fatih, P., Jianping, Y. (2018). Hyperparameter selection of one-class support vector machine by self-adaptive data shifting. *Pattern Recognition,* 74, pp. 198–211.

Wu, Y., Ngai, E. W. T., Wu, P., Wu, C. (2022). Fake news on the internet: a literature review, synthesis and directions for future research. *Internet research*, 32(5), pp. 1662–1699. DOI:10.1108/INTR-05-2021-0294.

Xu, R. (2021). Corporate fake news on social media. University of Miami Working paper.

Received: 12. 8. 2024
Reviewed: 2. 12. 2024
Accepted: 13. 12. 2024

Zuzana Janková, Ph.D.
Brno University of Technology
Faculty of Business and Management
Institute of Informatics
Kolejní 2906/4, 612 00 Brno
Czech Republic
E-mail: **Zuzana.Jankova@vutbr.cz**

Assoc. Prof. Michal Páleš, PhD.
University of Economics in Bratislava
Faculty of Economic Informatics
Department of Mathematics and Actuaries Sciences
Dolnozemská cesta 1, 852 35 Bratislava
Slovak Republic
E-mail: michal.pales@euba.sk

Marcin Potrykus, Ph.D.
Gdańsk University of Technology
Faculty of Management and Economics
Department of Finance
Gabriela Narutowicza 11/12, 80-233 Gdańsk
Poland
E-mail: marpotry@pg.edu.pl

Assoc. Prof. László Buics, Ph.D.
Széchenyi István University
Leadership & Mkt Department
Egyetem tér 1, 9026 Győr
Hungary
E-mail: buics.laszlo@sze.hu