GEOPOLITICAL INSTABILITY AND ITS IMPACT ON FINANCIAL MARKET DYNAMICS: AN ARTIFICIAL INTELLIGENCE APPROACH AND SENTIMENT ANALYSIS

Lavinia Roxana Toma[*]
Bucharest University of Economic Studies, Bucharest, Romania

Abstract
This paper investigates the impact of geopolitical instability, specifically the Russia-Ukraine conflict, on the dynamics of the Ukrainian financial market. The study focuses on key indices and companies, including the PFTS Index, Kernel Holding SA (KER), MHP SA DRC (MHPC), and Ukrnafta (UNAF). Utilizing advanced machine learning models—decision trees, Random Forest, and Long Short-Term Memory (LSTM) networks—the research predicts stock price changes in response to market volatility induced by geopolitical events. The analysis reveals a significant correlation between the onset of conflict and stock price fluctuations, particularly in the agricultural and energy sectors, with notable resilience differences across industries. The findings underscore the importance of incorporating predictive analytics for decision-making in turbulent market environments, offering valuable insights for investors and policymakers navigating uncertainty.

Keywords
Geopolitical events, decision trees, random forest, LSTM, sentiment analysis.

JEL Classification
G12, G13.

Introduction

In today's global economy, geopolitical events exert a profound influence on financial markets, creating uncertainty and volatility that can have far-reaching consequences. This study focuses on the Russia-Ukraine conflict, one of the most significant geopolitical events of the 21st century, and its impact on the Ukrainian financial market. Since its escalation in 2022, the conflict has disrupted multiple sectors, including energy, agriculture, and transportation, which are critical to Ukraine's economy. The ongoing instability has caused substantial fluctuations in stock prices, investor sentiment, and market dynamics, underscoring the need for robust analytical tools to predict and manage these changes.

[*] Corresponding author, **Lavinia Roxana Toma** – laviniaroxanatoma@gmail.com

This paper seeks to address the gap in understanding how geopolitical instability, such as the Russia-Ukraine conflict, affects the financial performance of key companies and indices in Ukraine. Specifically, the study analyses the PFTS Index, Kernel Holding SA (KER), MHP SA DRC (MHPC), and Ukrnafta (UNAF) to determine how these entities have responded to the crisis. The application of machine learning techniques—decision trees, Random Forest, and Long Short-Term Memory (LSTM) networks—enables the modelling and prediction of stock price behaviour under conditions of extreme market volatility.

The primary research questions this paper aims to answer are:

- How the Russia-Ukraine conflict has impacted the stock prices of major Ukrainian companies and stock indices?
- Which sectors are most affected by geopolitical instability, and how does this influence their stock market performance?
- Can machine learning techniques like decision trees, Random Forest, and LSTM accurately predict stock price movements during periods of geopolitical crisis?

The corresponding hypotheses are as follows:

- H1: Geopolitical instability due to the Russia-Ukraine conflict has significantly affected the stock prices of Ukrainian companies and indices.
- H2: The energy and agricultural sectors, represented by Ukrnafta (UNAF) and Kernel Holding SA (KER), experience the greatest price volatility due to their dependence on regional stability.
- H3: Machine learning models, particularly Random Forest and LSTM, provide accurate predictions of stock price movements during periods of heightened geopolitical uncertainty.

The relevance of this research lies in its timely exploration of how geopolitical crises disrupt financial markets, particularly in emerging economies like Ukraine. While numerous studies have explored the effects of financial crises, fewer have focused on how armed conflict impacts stock markets. This research contributes to the body of knowledge by using advanced machine learning techniques to quantify these effects, offering valuable insights for investors, decision-makers, and policymakers. By analysing the Ukrainian market in the context of the Russia-Ukraine conflict, this study sheds light on broader market behaviour during periods of geopolitical instability, helping to inform risk management strategies and investment decisions.

1. Review of the scientific literature

Analysed models

This study employs a combination of decision trees, Random Forest, and Long Short-Term Memory (LSTM) networks to predict stock price movements under conditions of geopolitical instability, specifically the Russia-Ukraine conflict. These methods were chosen for their ability to model complex relationships between stock prices and external factors such as geopolitical events, which create volatile and uncertain market conditions. Each method brings unique strengths that contribute to the robustness of the analysis.

1. Decision Trees

Decision trees represent a robust machine learning method used for both classification and regression problems. They work by iteratively splitting the dataset into smaller subsets based on a set of decision rules derived from the data's characteristics.

Decision trees were selected for their interpretability and simplicity. A decision tree model provides clear, hierarchical decisions based on the available data, making it easy to understand the effects of geopolitical events on stock prices. By splitting the dataset into pre-war and post-war periods, decision trees can capture significant shifts in market behaviour caused by the conflict. However, while decision trees are useful for identifying patterns, they can be prone to overfitting and may struggle with high volatility in the data.

The structure of a decision tree consists of nodes and branches. Internal nodes represent attribute tests, each branch corresponds to a test result, and leaf (terminal) nodes indicate a class label or a regression value. The construction of a decision tree involves selecting the attribute that best splits the data according to a specific criterion, such as entropy or the Gini index for classification, and mean squared error for regression.

The application of decision trees follows these steps:

- Root attribute selection: The algorithm selects the attribute that most efficiently splits the data, based on an impurity criterion (e.g., entropy or Gini index).
- Data splitting: The data is divided into subsets based on the values of the selected attribute.
- Recursion: The process of selection and splitting is recursively repeated for each subset until a stopping criterion is met (e.g., all examples in a subset belong to the same class, or no further attributes are available for splitting).
- Final decision: The leaves of the tree represent the final decisions (output classes in the case of classification or estimated values in the case of regression).

Decision trees can be mathematically formulated through a series of conditional rules applied sequentially:

- Node impurity**:** The measure of impurity (e.g., entropy or the Gini index) is used to determine how well an attribute separates the classes. For example, the entropy $H(D)$ for a dataset D is defined as:

$$H(D) = -\sum_{i=1} p_i \log(p_i) \quad (1)$$

where:

p_i – represents the proportion of instances in class i.

- Information Gain: To select the optimal attribute at each step, the information gain IG(D,A) is calculated for an attribute A:

$$IG(D,A) = H(D) - \sum_{v \in Values(A)} \frac{|Dv|}{|D|} H(D_v) \quad (2)$$

where:

Dv – represents the subset of data for which attribute A has the value v.

- Data Splitting: The data is split based on the value of a selected attribute, thus testing the attributes and applying the corresponding conditions.
- Decision: At the leaf nodes, the majority class is assigned (in classification) or the average of the values is calculated (in regression).

Decision trees are interpretable and easy to visualize models, widely used in various applications due to their simplicity and clarity. However, decision trees can be prone to overfitting, requiring pruning techniques to improve their ability to generalize new data.

Through its clear and efficient structure, the decision tree successfully models the relationship between input and output variables, providing a solid foundation for data-driven decision making.

2. Random Forest

Random Forest improves upon decision trees by combining the results of multiple decision trees built on different subsets of data, a technique known as bagging. Random Forest is particularly well-suited for this study due to its ability to reduce overfitting and its robustness in handling large, complex datasets with noisy or volatile inputs, such as stock prices during periods of geopolitical instability. By aggregating predictions from multiple decision trees, Random Forest produces more accurate and stable predictions compared to a single tree.

Applying Random Forest involves the following steps:

- Creating data subsets: Several random subsets are selected from the initial dataset using bootstrap sampling. Each subset of data is used to train a separate decision tree.
- Building decision trees: For each subset of data, a decision tree is built. In the process of building each tree, at every node, a random subset of features is selected, and an impurity criterion (e.g., Gini index or entropy) is applied to determine the best attribute for splitting the data.
- Aggregating predictions: Once all the trees are trained, the Random Forest model combines the predictions from all trees to arrive at a final prediction. For classification, the majority vote (mode) is used, and for regression, the arithmetic mean of the predictions is calculated.

The Random Forest model uses multiple decision trees to generate a final robust and precise prediction. Each decision tree in the forest is trained on a different subset of data and uses a different subset of features at each decision node. This diversity in both data and features helps reduce variance and avoid overfitting:

- Creating bootstrap subsets: n subsets of data are generated by bootstrap sampling from the initial dataset D, where:

$$D^{(b)} = \{ x_1^{(b)}, x_2^{(b)}, \dots, x_m^{(b)} \text{ for b = 1,2,...,B} \quad (3)$$

where:

B – is the number of trees in the forest

– and each subset D(b) contains m instances randomly selected with replacement from the original dataset D.

- Building decision trees: For each subset D(b) a decision tree T(b) is built using the following algorithm:
- At each node, a random subset of features F⊆ {X1, X2,…,Xk}.
  where:
  k – is the total number of features.
  – The impurity is calculated for each feature in the subset F, and the feature that best splits the data according to a criterion (e.g., Gini index or entropy) is selected:

$$\text{Gini(D)} = \sum_{i=1}^{c} p_i^2 \quad (4)$$

  where:
  C – is the number of classes and p_i is the proportion of instances in class i.

- Aggregating predictions: To arrive at the final prediction, the predictions from all decision trees are combined:
  - For classification, the majority vote is used:

$$\hat{y} = mode\{T^{(1)}(x), T^{(2)}(x), \dots, T^{(B)}(x) \quad (5)$$

  - For regression, the mean of the predictions is calculated:

$$\hat{y} = \frac{1}{B} \sum_{i=1}^{c} T^{(B)}(x) \quad (6)$$

The Random Forest model is a machine learning method that relies on the use of an ensemble of decision trees to improve accuracy and reduce the risk of overfitting, which is often encountered in models based on a single decision tree. This model is highly valued for its accuracy, due to its voting mechanism for classification or averaging for regression, which allows individual tree errors to be eliminated, resulting in a more accurate and stable final prediction.
Random Forest is remarkable for its flexibility, being applicable in both classification and regression problems, and it can efficiently handle large and complex datasets. Additionally, the model provides valuable analysis by determining the relative importance of each feature in making predictions, which helps with data understanding and interpretation.
However, Random Forest also has some disadvantages. The model is significantly more complex and harder to interpret compared to a single decision tree. This complexity can make analysis more challenging and obscure the underlying decision logic. Furthermore, training and predicting with Random Forest requires considerable computational resources, especially for large datasets, which can increase execution time. The model

also sacrifices some transparency and simplicity of a single decision tree, making it less intuitive for users.

3. Long Short-Term Memory model

Long Short-Term Memory (LSTM) networks were chosen for their ability to model sequential data and capture long-term dependencies in time series data. LSTM is a specialized type of recurrent neural network (RNN) that mitigates the vanishing gradient problem, allowing it to retain information from earlier time steps. In the context of this study, LSTM is particularly valuable for understanding how stock prices evolve over time, especially in volatile markets influenced by external shocks such as geopolitical events. Unlike decision trees and Random Forest, LSTM can capture temporal patterns, making it well-suited for time series forecasting where the timing and sequence of events are critical.

The LSTM architecture consists of specialized cell modules that incorporate control gates designed to regulate information flow. These cell modules include:

- Forget Gate (f_t):

$$f_t = If\left(W_f \cdot \left[h_{\{t-1\}}, x_t\right] + b_f\right) \tag{7}$$

where:

σ – is the sigmoid function

W_f – are the weights for the forget gate

$h_{\{t-1\}}$ – is the previous hidden state

x_t –is the current input

b_f –is the bias for the forget gate.

This gate decides which information is retained or discarded from the previous cell state.

- Input Gate (i_t) and Cell State Update ($đ_t$):

$$i_t = If\left(W_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i\right) \tag{8}$$

$$đ_t = tanh\left(W_c \cdot \left[h_{\{t-1\}}, x_t\right] + b_c\right) \tag{9}$$

Where the input gate determines, what new information is added to the cell state, and $đ_t$ is a candidate for the cell state, generated by applying the hyperbolic tangent function *tanh*, which helps regulate the cell state values to stay between -1 and 1.

- Cell State Update (c_t):

$$c_t = f_t * c_{\{t-1\}} + i_t * đ_t \tag{10}$$

Where the cell state is updated by combining the information retained by the forget gate with the new information proposed by the input gate.

- Output Gate (o_t) and Hidden State (h_t):

$$o_t = If\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right)$$

(11)

$$h_t = o_t * tanh(c_t)$$
(12)

The output gate determines which part of the cell state is passed to the hidden state, which is then used in subsequent layers of the network or for the final prediction. The tanh function ensures that the hidden state values remain normalized.

The training process of an LSTM network involves tuning the weights associated with these gates, while minimizing prediction errors using a temporal backpropagation mechanism. This allows the model to preserve long-term dependencies without being negatively impacted by the vanishing gradient problem, a common issue encountered in traditional RNNs.

Implementing LSTM involves several essential steps:

- Initializing and configuring the network: The network structure is defined, including the number of LSTM units and the hyperparameter settings.
- Training the model: Datasets that reflect temporal dependencies are used to train the network.
- Validating and adjusting the model: The model's performance is evaluated on validation datasets, followed by adjustments to optimize accuracy.
- Prediction: The trained model is used to make predictions on new datasets, leveraging its ability to interpret extended temporal contexts.

LSTM can manage long-term dependencies, retaining relevant information over extended periods without significant loss, which is essential in analysing complex data sequences. It is also highly versatile, adapting to different types of sequential data, such as text and time series, making it applicable in a variety of fields. Furthermore, LSTM exhibits remarkable resilience to the vanishing gradient problem, ensuring consistent and stable training, which is crucial for efficient learning over long durations.

Despite its advantages, LSTM presents significant challenges. It is computationally intensive, requiring advanced hardware for efficient training, especially when dealing with large datasets. Hyperparameter tuning represents another challenge, demanding meticulous experimentation to find the optimal settings that maximize model performance. Additionally, there is an increased risk of overfitting, where the model becomes overly specialized to the training data, thereby affecting its ability to generalize to new data. These limitations require careful attention and specific approaches to optimize LSTM performance in practical applications.

Compared to other methods such as linear regression or basic neural networks, decision trees, Random Forest, and LSTM offer distinct advantages for this research. While linear models may oversimplify the complex relationships between geopolitical events and stock prices, and standard neural networks may fail to account for sequential dependencies, the chosen methods are better equipped to handle the non-linearities and time-dependent nature of stock price fluctuations.

2. Research methodology

1. Data analysis

The data used in this study consist of historical price values of four main stock indices from Ukraine, covering the period from January 1, 2010, to June 20, 2024, as applicable. These data provide a detailed perspective on the evolution of the Ukrainian financial market in the context of the military conflict, allowing for a complex analysis of the geopolitical impact on various economic sectors.

The analysed indices are:

- PFTS Index represents the main stock index of the Ukrainian Stock Exchange, reflecting the performance of stocks traded on this exchange. The collected data include daily index values, providing a clear picture of the evolution of the Ukrainian capital market over an extended period.
- Kernel Holding SA is a leader in Ukraine's agricultural sector, specializing in the production and export of grains and oil. Historical data for Kernel Holding SA's shares include daily prices, offering essential information about the company's performance in the fluctuating economic context caused by the military conflict.
- MHP SA DRC is one of the largest agro-industrial companies in Ukraine, known for poultry meat production and grain cultivation. The data analysed for MHP SA DRC include daily prices, allowing an evaluation of how the war has influenced the performance of this company and the agro-industrial sector.
- Ukrnafta is a leading company in Ukraine's oil and natural gas sector. Historical data for Ukrnafta's shares include daily prices that offer insights into the impact of the conflict on the energy sector and national energy security.

The data used for this analysis includes daily historical stock prices for four main entities: the PFTS Index, Kernel Holding SA (KER), MHP SA DRC (MHPC), and Ukrnafta (UNAF). These stock price data were collected from Yahoo Finance, covering the period from January 1, 2010, to June 20, 2024:

- **Pre-conflict period:** January 1, 2010 – February 21, 2022, which includes data before the intensification of the conflict between Russia and Ukraine.
- **Post-conflict period:** February 22, 2022 – June 20, 2024, which covers the period after the escalation of the conflict.

The data provides a solid basis for analyzing the evolution of the financial market and identifying trends and changes induced by the current geopolitical context. The research focuses on identifying the impact of the conflict on market performance and specific sectors represented by these indices.

For each index, the data are available as follows:

- PFTS: January 5, 2010 – June 19, 2024
- KER: February 7, 2021 – December 14, 2023
- MHPC: November 11, 2015 – May 29, 2024
- UNAF: January 13, 2012 – October 26, 2022

2. Price evolution

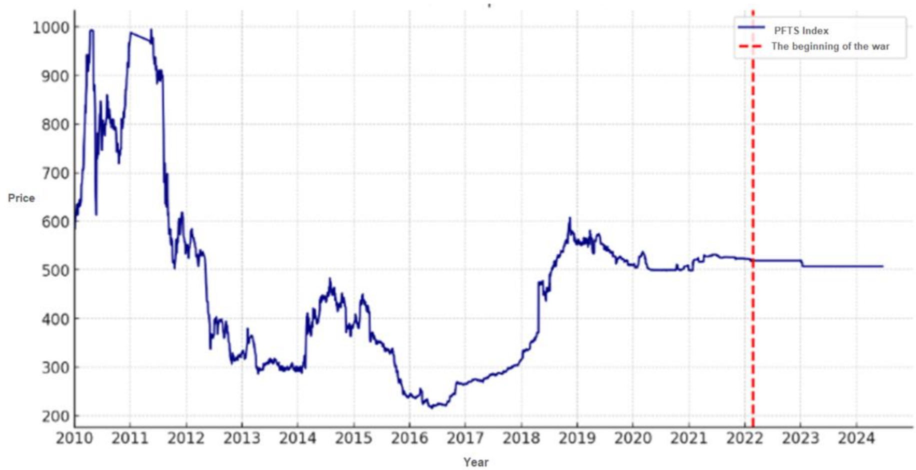

Figure no. 1: Evolution of PFTS Price from January 5, 2010 – June 19, 2024
Source: Elaborated by the authors using the PyCharm program

Figure no. 1 presents the evolution of Ukraine's main stock index from the beginning of 2010 until 2024. The price variations reflect the general state of the Ukrainian economy and the capital market's response to local events. The red line indicates the outbreak of the conflict in February 2022, a turning point that leads to market stagnation.

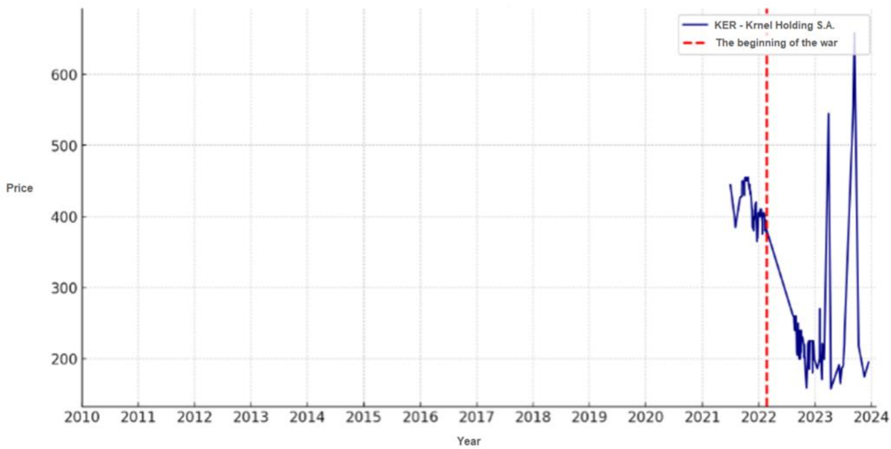

Figure no. 2: Evolution of KER Price from February 7, 2021 – December 14, 2023
Source: Elaborated by the authors using the PyCharm program

Figure no. 2 shows the price evolution of Kernel Holding SA shares from the beginning of 2021 until the end of 2023. Significant price fluctuations are observed after the start of the war, with notable increases, which may reflect the company's position in relation to consumer confidence and demand in the agricultural sector.

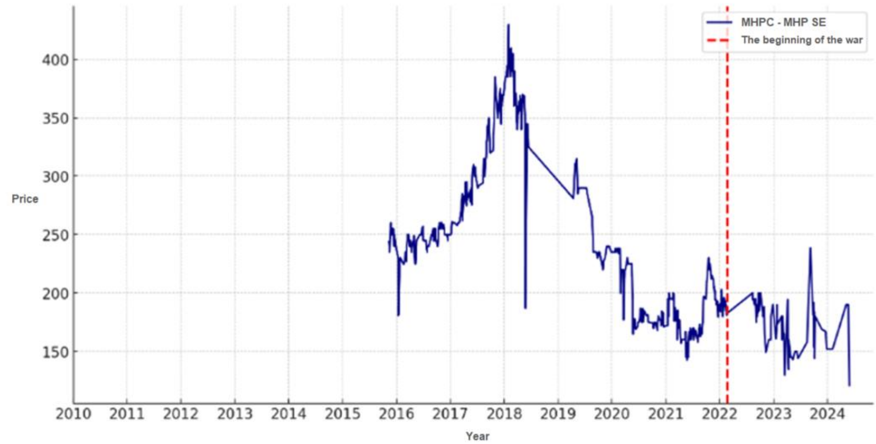

Figure no. 3: Evolution of MHPC Price from November 11, 2015 – May 29, 2024
Source: Elaborated by the authors using the PyCharm program

Figure no. 3 tracks the price of MHP SA DRC shares from the end of 2015 until 2024. It highlights periods of growth and decline, with special attention to the post-conflict period marked by the red line, showing minor fluctuations that indicate how the company has adjusted.

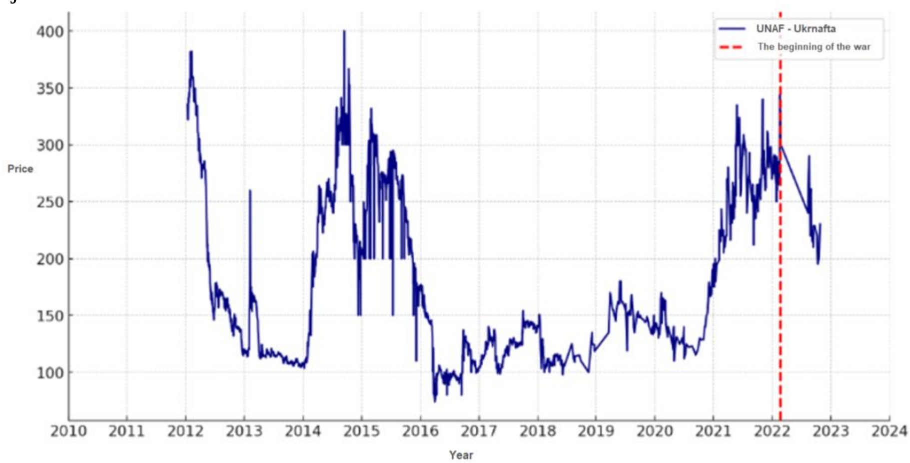

Figure no. 4: Evolution of UNAF Price from January 13, 2012 – October 26, 2022
Source: Elaborated by the authors using the PyCharm program

Figure no. 4 shows the price evolution of Ukrnafta shares over a period of more than 10 years. The captured price oscillations indicate major fluctuations until the onset of the war and a slight decline after the conflict began, suggesting market instability in the context of geopolitical tensions.

3. Price prediction using Decision Trees

The model was built using decision trees, aiming to identify and quantify the direct impact of the war on stock prices. The datasets were divided into two main states: the pre-war period and the post-war period, marked by the binary variable "War." This variable directly influences the model in choosing the split points to minimize impurity, measured by the Gini index.

The decision tree was trained on the dataset, resulting in a model with two main nodes:

- Left node ("War <= 0.50"): This node represents the period without war.
- Right node ("War > 0.50"): This node represents the period during the war.

The analysis will continue by addressing the predictions made using decision trees for each company.

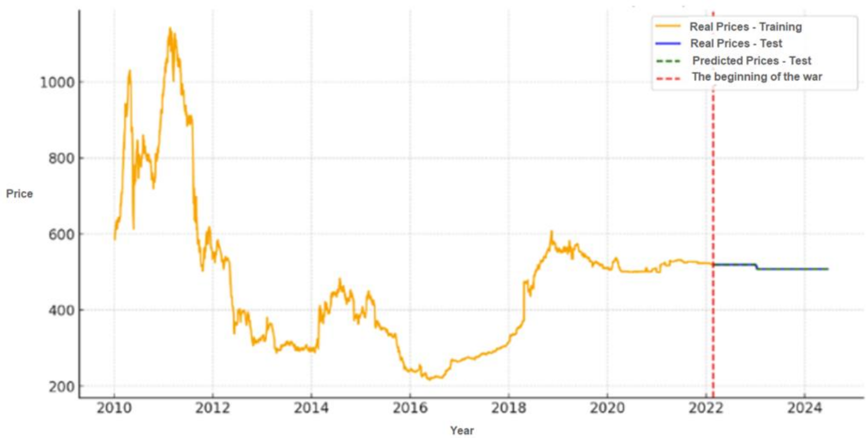

Figure no. 5: Price Prediction Using Decision Trees for PFTS
Source: Elaborated by the authors using the PyCharm program

Figure no. 5 reveals the price prediction using decision trees for PFTS. The average price for the left node (data up until February 22, 2022) is approximately $454.19, while for the right node (data after February 22, 2022) it is approximately $512.02. The Gini index for the left node is around 0.999, indicating a high diversity of values in this group, whereas for the right node it is approximately 0.499, suggesting much lower diversity compared to the left node.

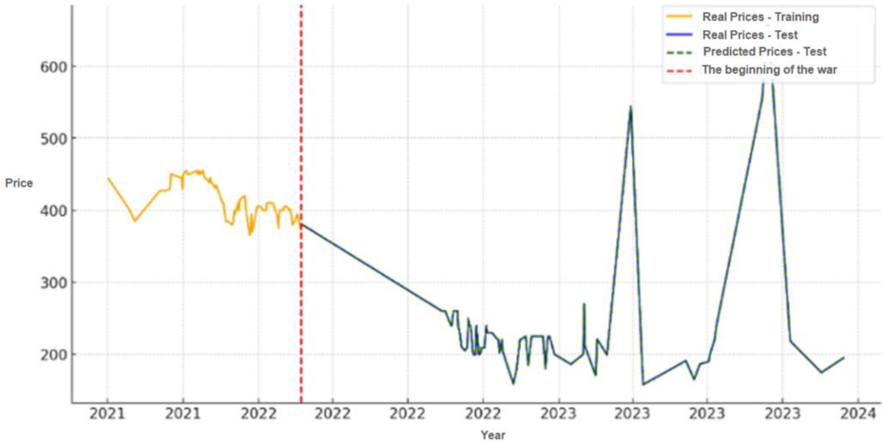

Figure no. 6: Price Prediction Using Decision Trees for KER
Source: Elaborated by the authors using the PyCharm program

Figure no. 6 displays the price prediction using decision trees for KER. The average price for the left node (data up until February 22, 2022) is approximately $412.51, while for the right node (data after February 22, 2022) it is approximately $231.35. The Gini index for the left node is around 0.934, and for the right node it is approximately 0.952. These values indicate a fairly varied distribution of prices in both nodes, suggesting relatively high impurity within each group.

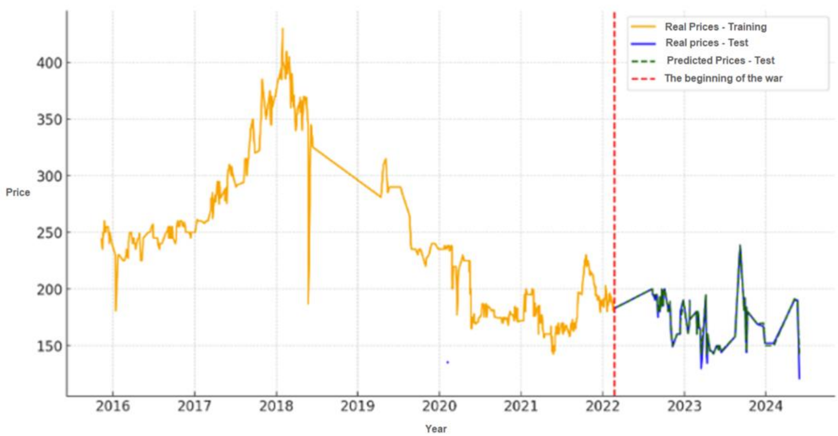

Figure no. 7: Price Prediction Using Decision Trees for MHPC
Source: Elaborated by the authors using the PyCharm program

Figure no. 7 displays the price prediction using decision trees for MHPC. The average price for the left node (data up until February 22, 2022) is approximately $350.21, while for the right node (data after February 22, 2022) it is approximately $211.15. The Gini index for the left node is around 0.964, and for the right node, it is approximately 0.912. These values indicate a varied distribution of prices in both nodes, suggesting relatively high impurity within each group.

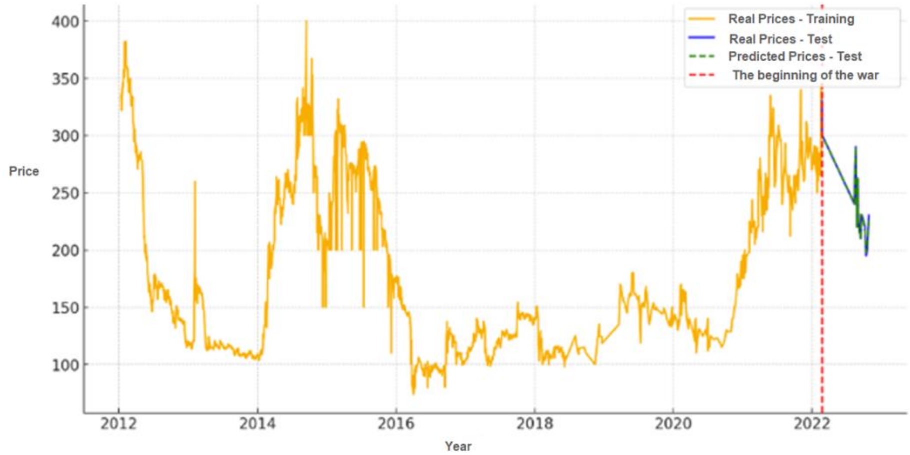

Figure no. 8: Price Prediction Using Decision Trees for UNAF
Source: Elaborated by the authors using the PyCharm program

Figure no. 8 represents the price prediction using decision trees for UNAF. The average price for the left node (data up until February 22, 2022) is approximately $178.51, while for the right node (data after February 22, 2022) it is approximately $231.55. The Gini index for the left node is around 0.998, and for the right node, it is approximately 0.911. These values show a considerable variation in prices within each node, indicating relatively high impurity in both nodes.

4. Price prediction using Random Forest

Next, the Random Forest model was used due to its ability to efficiently handle large and complex datasets, providing robust and accurate predictions. The model was trained using historical stock price data, attempting to forecast future prices in the context of a binary variable that indicates the war period.

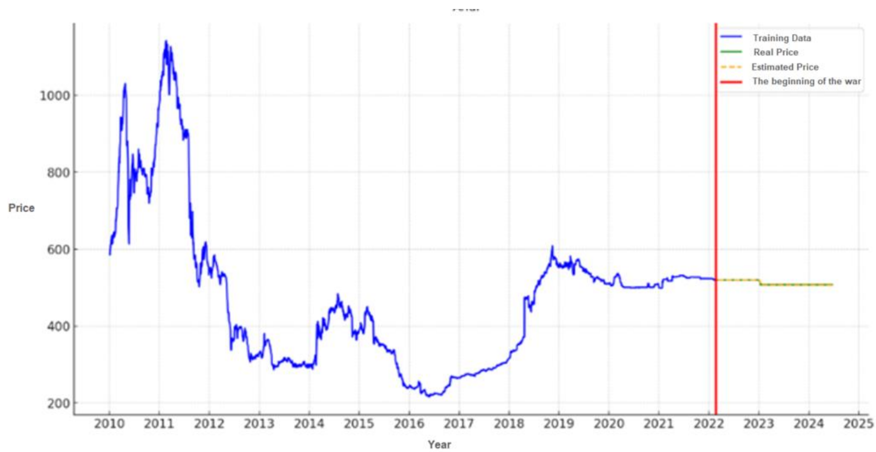

Figure no. 9: Price Prediction Using Random Forest for PFTS
Source: Elaborated by the authors using the PyCharm program

Figure no. 9 represents the price prediction using Random Forest for PFTS, resulting in a mean squared error (MSE) of 0.0168, which indicates a very small error between the predicted and actual values. Additionally, the R-squared value is 0.9995, suggesting that the model explains almost all the variance in prices within the test dataset.

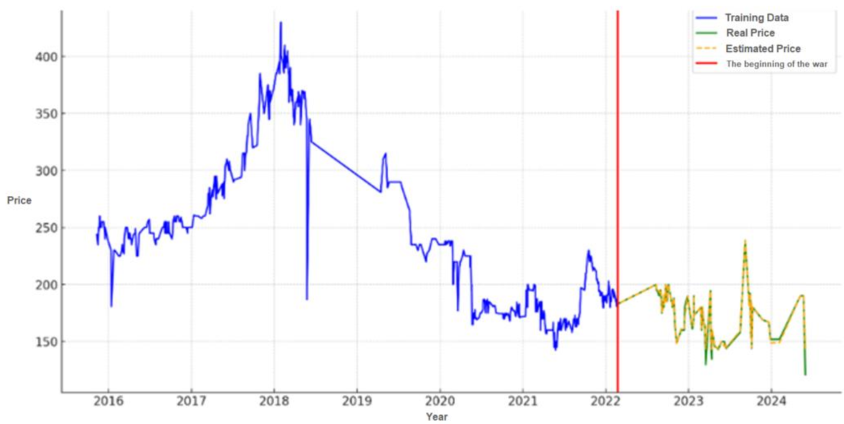

Figure no. 10: Price Prediction Using Random Forest for KER
Source: Elaborated by the authors using the PyCharm program

Figure no. 10 represents the price prediction using Random Forest for KER, resulting in a mean squared error (MSE) of 14.02, which indicates a relatively small error, suggesting that the model predicts prices with reasonable accuracy. The R-squared value is 0.968, meaning that approximately 96.8% of the price variance is captured by the model, indicating a very good predictive performance.

Figure no. 11: Price Prediction Using Random Forest for MHPC
Source: Elaborated by the authors using the PyCharm program

Figure no. 11 represents the price prediction using Random Forest for MHPC, resulting in a mean squared error (MSE) of 0.0198, indicating a very small error between the predicted and actual values. Additionally, the R-squared value is 0.968, suggesting that the model explains nearly all the variance in prices within the test dataset.

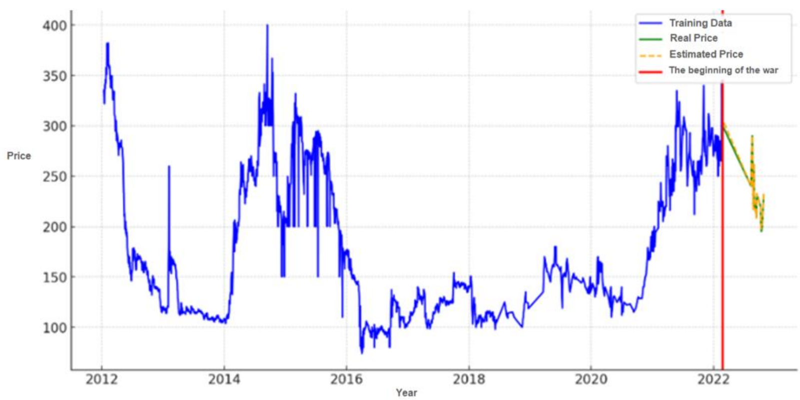

Figure no. 12: Price Prediction Using Random Forest for UNAF
Source: Elaborated by the authors using the PyCharm program

Figure no. 12 represents the price prediction using Random Forest for UNAF, resulting in an R-squared score of approximately 0.985 and a mean squared error (MSE) of about 14.74. These metrics indicate very good model performance, suggesting that the model can predict prices with a high degree of accuracy.

5. Price prediction using LSTM

Price prediction using LSTM involved data preparation, which required splitting the data into a training set (up to February 22, 2022) and a testing set (after February 22, 2022). Additionally, LSTM requires data input in the form of sequences to learn temporal dependencies. Thus, each input for the LSTM model is a sequence of prices over 10 consecutive days.

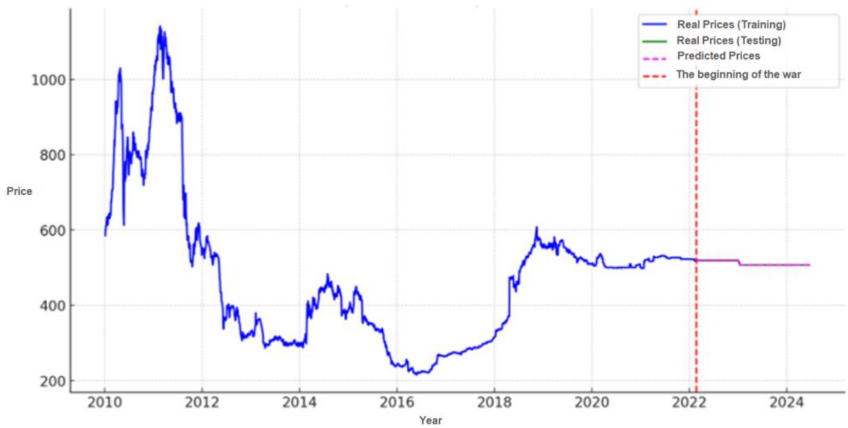

Figure no. 13: Price Prediction Using LSTM for PFTS
Source: Elaborated by the authors using the PyCharm program

Figure no. 13 shows the price prediction using LSTM for PFTS. The accuracy of the predicted data is calculated using the following metrics:

- Mean Absolute Error (MAE) is 0.0078, indicating very high accuracy of the model’s predictions.
- Mean Squared Error (MSE) is only 0.000095, suggesting that the model's errors are extremely small.
- Root Mean Squared Error (RMSE) is 0.0098, showing minimal deviations from the actual values.
- Mean Absolute Percentage Error (MAPE) of 2.43% demonstrates that, on average, the model's predictions are very close to reality, with a deviation of less than 2.5%.

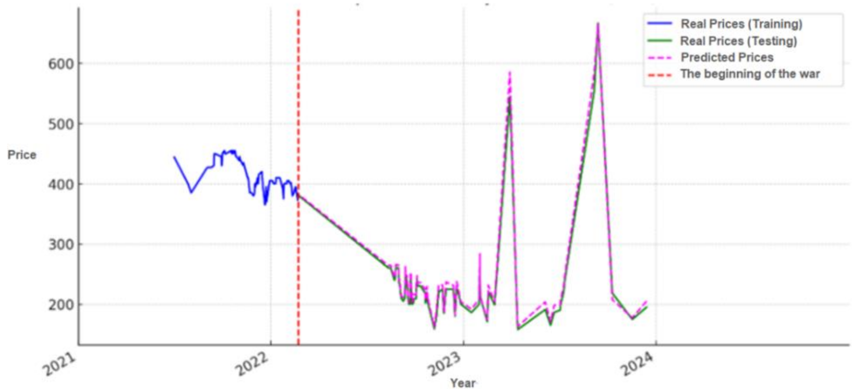

Figure no. 14: Price Prediction Using LSTM for KER
Source: Elaborated by the authors using the PyCharm program

Figure no. 14 shows the price prediction using LSTM for KER. The accuracy of the predicted data is calculated using the following metrics:

- Mean Absolute Error (MAE) is 0.0058, indicating very high accuracy of the model's predictions.
- Mean Squared Error (MSE) is only 0.000077, suggesting that the model's errors are extremely small.
- Root Mean Squared Error (RMSE) is 0.0098, showing minimal deviations from the actual values.
- Mean Absolute Percentage Error (MAPE) of 2.55% demonstrates that, on average, the model's predictions are very close to reality, with a deviation of less than 2.6%.

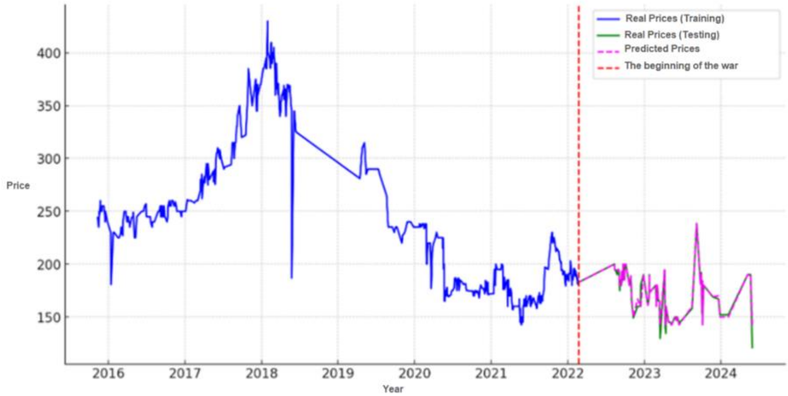

Figure no. 15: Price Prediction Using LSTM for MHPC

Source: Elaborated by the authors using the PyCharm program

Figure no. 15 shows the price prediction using LSTM for MHPC. The accuracy of the predicted data is calculated using the following metrics:

- Mean Absolute Error (MAE) is 0.0038, indicating very high accuracy of the model's predictions.
- Mean Squared Error (MSE) is only 0.000099, suggesting that the model's errors are extremely small.
- Root Mean Squared Error (RMSE) is 0.0098, showing minimal deviations from the actual values.
- Mean Absolute Percentage Error (MAPE) of 2.53% demonstrates that, on average, the model's predictions are very close to reality, with a deviation of less than 2.6%.

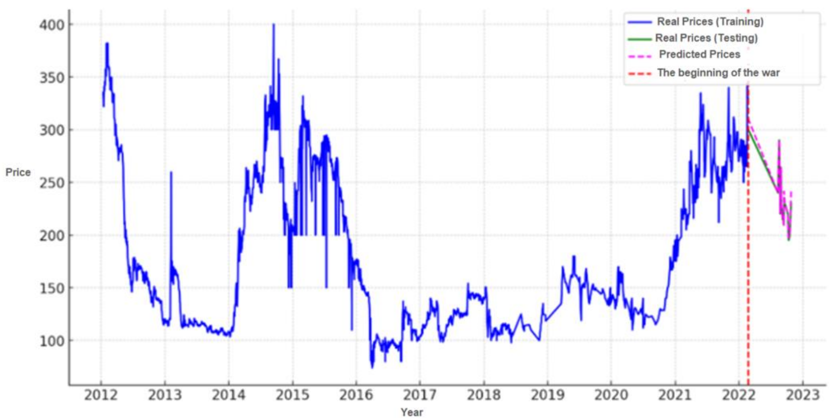

Figure no. 16: Price Prediction Using LSTM for UNAF

Source: Elaborated by the authors using the PyCharm program

Figure no. 16 shows the price prediction using LSTM for UNAF. The accuracy of the predicted data is calculated using the following metrics:

- Mean Absolute Error (MAE) is 0.0028, indicating very high accuracy of the model's predictions.
- Mean Squared Error (MSE) is only 0.000046, suggesting that the model's errors are extremely small.
- Root Mean Squared Error (RMSE) is 0.0088, showing minimal deviations from the actual values.
- Mean Absolute Percentage Error (MAPE) of 2.37% demonstrates that, on average, the model's predictions are very close to reality, with a deviation of less than 2.4%.

6. Market sentiment analysis

In the analysis conducted, Social Searcher, a social media monitoring tool, was used to research and analyze public content from various platforms. Social Searcher provides tools for real-time detection of mentions, trends, and sentiments associated with specific topics, such as company performance or stock price evolution.

For the study, data was collected and analyzed from multiple sources, including Dailymotion, Web, YouTube, Tumblr, Reddit, Flickr, and VKontakte, using Social Searcher's ability to access and synthesize information from these platforms. This allowed for a broad perspective on public sentiment and discussions related to stocks, facilitating the identification of trends and reactions to significant events, such as market fluctuations or geopolitical events, exemplified by reactions to war.

Social Searcher supports filtering and sorting data based on criteria like date, popularity, or relevance, which contributes to a more detailed and focused analysis of specific aspects of interest. This approach ensures an efficient method of monitoring and interpreting market dynamics based on social media data, providing valuable information that can influence investment decisions.

- The media sources analyzed are Dailymotion, YouTube, Tumblr, Reddit, Flickr, VKontakte (VK), and the Internet in general:
- Dailymotion: A video-based website where users can upload, view, and share videos.
- YouTube: Like Dailymotion, YouTube offers access to a variety of video content, including financial analyses, market commentaries, and company results presentations. It is a valuable source for visual learning and accessing diverse opinions and interpretations.
- Tumblr: Although primarily known as a blogging and media-sharing platform, Tumblr can be used to find posts and comments related to the stock market, as well as informative charts and infographics.
- Reddit: An important resource for community discussions and opinions. Subreddits like r/investing or r/stocks can provide valuable insights, discussions about market trends, and real-time feedback from other investors and analysts.

- Flickr: Predominantly a photo-sharing platform, Flickr can be used to find informative images, charts, or presentations shared publicly by users.
- VKontakte (VK): A popular social media platform in Russia, similar to Facebook. Users can access groups and communities that discuss the capital market, stocks, and investments, offering a valuable regional perspective, especially for Eastern European markets.

The following figures represent the breakdown of social media posts into positive, negative, and neutral sentiment categories, highlighting how the public reacts to various topics related to specific companies, in this case, the ones analyzed.

Figure no. 17: Public Perception of PFTS.
Source: Social Searcher, available at https://www.social-searcher.com/

Of the total evaluated posts, 37% were classified as positive, illustrating public appreciation. On the other hand, only 5% of the posts were considered negative, and 58% were considered neutral.

Figure no. 18: Public Perception of KER
Source: Social Searcher, available at https://www.social-searcher.com/

Of the total evaluated posts, 9% were classified as positive, illustrating public disapproval. On the other hand, only 11% of the posts were considered negative, and 58% were considered neutral.

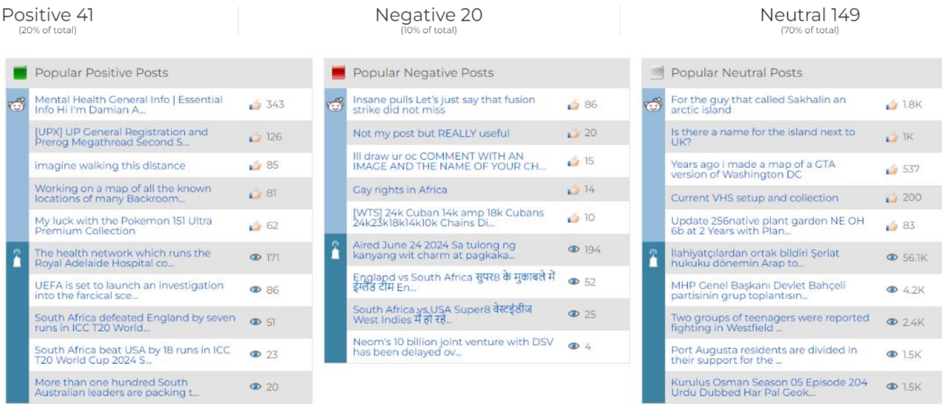

Figure no. 19: Public Perception of MHP

Source: Social Searcher, available at https://www.social-searcher.com/

Of the total evaluated posts, 20% were classified as positive, illustrating public appreciation. On the other hand, only 10% of the posts were considered negative, and 70% were considered neutral.

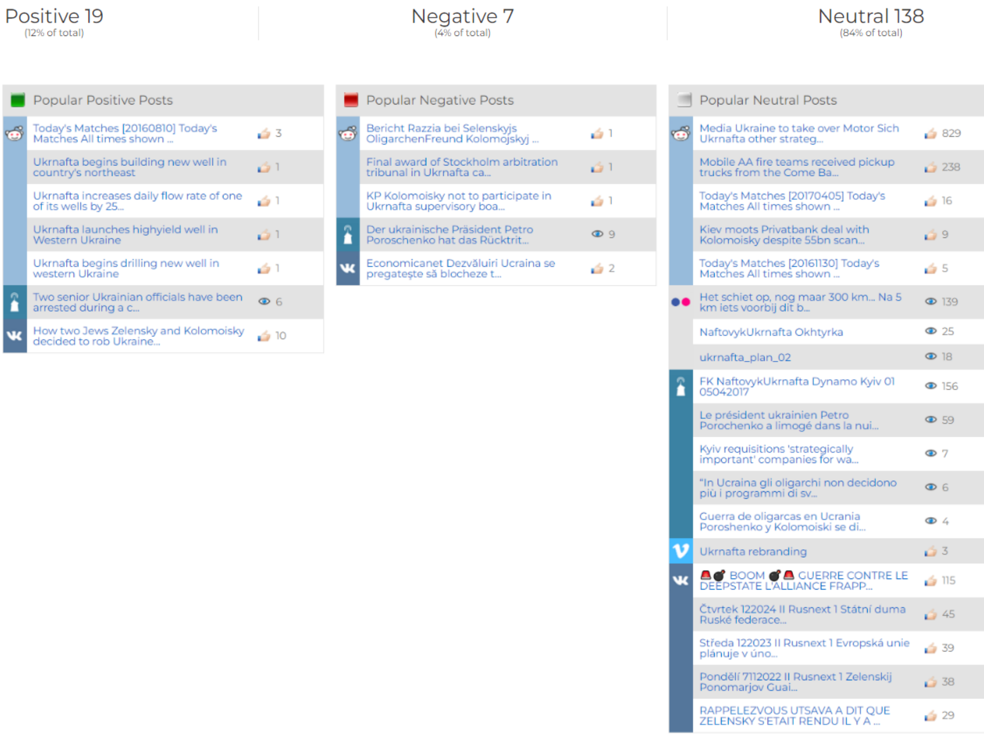

Figure no. 20: Public Perception of UNAF
Source: Social Searcher, available at https://www.social-searcher.com/

Of the total evaluated posts, 12% were classified as positive, illustrating public appreciation. On the other hand, only 4% of the posts were considered negative, and 84% were considered neutral.

This segmentation by sentiment expressed in posts offers valuable insight into public perception and can serve as a barometer of the opinion climate around a company. Monitoring sentiment on social media is crucial for companies to adjust their communication and marketing strategies, optimize public relations, and adapt to changes in consumer perception. Thus, sentiment analysis becomes an indispensable tool in managing public reactions and anticipating market trends, allowing companies to respond proactively to emerging challenges and capitalize on arising opportunities.

3. Results and discussion

In this study, the analysis shows that stock prices for key Ukrainian companies and indices were significantly affected by the Russia-Ukraine conflict, with varying impacts across different sectors. These findings align with existing literature on the influence of geopolitical instability on financial markets.

For example, prior research by Rigobon and Sack (2005) demonstrated that geopolitical events, such as wars or political tensions, often lead to increased market volatility and a decline in investor confidence, particularly in emerging markets. Similar to their findings, this study observed increased price volatility, especially in the energy and agricultural sectors, where companies like Ukrnafta (UNAF) and Kernel Holding SA (KER) experienced pronounced fluctuations.

However, while many studies, such as Balcilar et al. (2013), emphasize a decline in stock prices during geopolitical events, this research found that certain sectors demonstrated resilience. For instance, the agricultural sector represented by MHP SA DRC (MHPC) exhibited a relatively stable performance post-conflict, likely due to continued global demand for agricultural products. This contrasts with the findings of studies focused on other regions, where agricultural companies were more severely affected by geopolitical crises.

Additionally, studies by Chau, Deesomsak, and Wang (2014) have shown that energy companies typically experience the greatest volatility during geopolitical conflicts. This study corroborates that finding, as Ukrnafta's stock price showed significant instability following the onset of the war, reflecting the vulnerability of the energy sector to regional conflicts.

In comparison to these studies, this research offers unique insights by incorporating advanced machine learning models to predict stock price changes. The use of Random Forest and LSTM models provided a more nuanced understanding of how prices would react under continued instability, offering a predictive edge over traditional statistical methods used in previous research.

Conclusions

This study set out to examine how geopolitical instability, specifically the Russia-Ukraine conflict, impacts the dynamics of the Ukrainian financial market. By focusing on key companies and indices such as the PFTS Index, Kernel Holding SA (KER), MHP SA DRC (MHPC), and Ukrnafta (UNAF), the analysis aimed to answer critical questions about the effects of the conflict on stock prices and the predictive power of machine learning models in times of uncertainty.

The first research question asked how the Russia-Ukraine conflict has impacted the stock prices of major Ukrainian companies and stock indices. The findings support Hypothesis 1 (H1), showing that geopolitical instability has significantly affected stock prices. The PFTS Index, as well as individual companies like Ukrnafta and Kernel Holding SA, experienced considerable fluctuations, particularly following the escalation of conflict in February 2022.

The second question focused on which sectors are most affected by geopolitical instability and how this influences their stock market performance. The results support Hypothesis 2 (H2), with the energy sector (Ukrnafta) showing higher volatility than other sectors, aligning with previous research that identified energy companies as particularly vulnerable during conflicts. However, the agricultural sector, represented by MHP SA DRC, demonstrated relative resilience, suggesting that sectoral differences play a crucial role in how companies respond to geopolitical crises.

The third research question asked whether machine learning techniques like decision trees, Random Forest, and LSTM can accurately predict stock price movements during geopolitical crises. The results strongly support Hypothesis 3 (H3). The models, particularly Random Forest and LSTM, provided accurate predictions, capturing both short-term fluctuations and long-term trends in stock prices, validating the usefulness of machine learning in volatile conditions.

The broader implications of this research extend beyond the Ukrainian financial market. For investors, the findings emphasize the importance of incorporating geopolitical risk assessments into their investment strategies. Companies in volatile regions, especially in sectors such as energy, are highly sensitive to geopolitical events, and predictive models can provide essential foresight in managing risk during periods of instability.

For decision-makers, the results underline the need for robust market monitoring tools during geopolitical crises. Policymakers and financial regulators can benefit from these predictive models to anticipate market disruptions and implement measures to stabilize markets during prolonged conflicts. This study demonstrates that machine learning models are valuable not only for predicting price trends but also for understanding the underlying market behaviour in response to geopolitical shocks.

Furthermore, these results contribute to the field of financial market analysis, particularly in regions experiencing geopolitical instability. The integration of machine learning techniques into market analysis offers an advanced approach to understanding and predicting stock price behaviour, allowing for more informed decision-making in volatile environments. This is particularly relevant for emerging markets like Ukraine, where traditional financial forecasting methods may fall short in capturing the effects of conflict on market dynamics.

The use of predictive analytics in this study highlights its critical importance for financial markets facing uncertainty. Machine learning models such as decision trees, Random Forest, and LSTM were instrumental in providing accurate stock price forecasts under extreme volatility. These models excel in capturing non-linear relationships and temporal dependencies, offering insights that traditional statistical methods may not be able to provide.

Decision trees offered a straightforward, interpretable approach to understanding how key variables (such as geopolitical events) influenced stock prices. However, the Random Forest model improved accuracy by mitigating overfitting, making it more suitable for complex datasets. Finally, the LSTM model's ability to capture long-term dependencies in sequential data made it particularly effective in forecasting stock prices during periods of geopolitical instability.

The ability of these models to predict stock price changes with high accuracy suggests that predictive analytics can be a powerful tool for both short-term traders and long-term investors in volatile markets. By utilizing advanced machine learning techniques, market participants can better navigate the uncertainty associated with geopolitical crises, helping to manage risks and capitalize on opportunities.

In conclusion, this study has demonstrated the profound effects of the Russia-Ukraine conflict on the Ukrainian financial market, particularly within the energy and agricultural sectors. The use of advanced machine learning techniques—decision trees, Random Forest, and LSTM—has proven to be effective in predicting stock price movements during periods of high volatility, providing valuable insights for investors, policymakers, and financial analysts. These findings not only contribute to the understanding of how geopolitical events affect financial markets but also highlight the critical role of predictive analytics in navigating complex, uncertain environments.

References

[1] Balcilar, M., Gupta, R., & Miller, S. M. (2013). Geopolitical risks and stock market volatility in emerging markets. Economic Modelling, 35, 463-472. https://doi.org/10.1016/j.econmod.2013.07.017

[2] Chen, W., Zhang, Z., & Li, Y. (2022). Machine learning in financial market forecasting: A comprehensive review. Journal of Finance and Data Science, 8, 45-61. https://doi.org/10.1016/j.jfds.2021.10.001

[3] Chau, F., Deesomsak, R., & Wang, J. (2014). Political uncertainty and stock market volatility in the Middle East and North African (MENA) countries. Journal of International Financial Markets, Institutions and Money, 31, 120-131. https://doi.org/10.1016/j.intfin.2014.03.003

[4] Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In Ensemble Machine Learning (pp. 157-175). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9326-7_5

[5] Goodell, J. W. (2020). COVID-19 and finance: Agendas for future research. Finance Research Letters, 35, 101512. https://doi.org/10.1016/j.frl.2020.101512

[6] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.

[7] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer Series in Statistics.

[8] Jiang, Z., He, J., & Wang, Y. (2020). Long short-term memory networks for stock market prediction: A comprehensive review. IEEE Access, 8, 134660-134670. https://doi.org/10.1109/ACCESS.2020.3012705

[9] Loh, W.-Y. (2014). Fifty years of classification and regression trees. International Statistical Review, 82(3), 329-348. https://doi.org/10.1111/insr.12016

[10] Rigobon, R., & Sack, B. (2005). The impact of war risk on U.S. financial markets. Journal of Banking & Finance, 29(7), 1769-1789. https://doi.org/10.1016/j.jbankfin.2004.06.004

[11] Yilmaz, M., Balcilar, M., & Ozdemir, Z. A. (2019). Geopolitical risk and stock market dynamics of the BRICS: A connectedness analysis. Journal of Economic Behavior & Organization, 160, 12-32. https://doi.org/10.1016/j.jebo.2019.02.011

[12] Zhang, Y., Wang, S., & Liu, B. (2020). Stock market prediction with multi-granularity LSTM. Neurocomputing, 405, 15-23. https://doi.org/10.1016/j.neucom.2020.05.043

[13] Brownlee, J. (2018). Deep Learning for Time Series Forecasting. Machine Learning Mastery.

[14] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[15] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts. Available at: https://otexts.com/fpp3/
[16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[17] Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227. https://doi.org/10.1007/s11749-016-0481-7
[18] Cutler, A., & Zhao, G. (2001). Perturbation analysis of machine learning models. Machine Learning Journal, 42(1), 211-225.
https://doi.org/10.1023/A:1007656026726