Journal of Research Administration
Society of Research Administrators International

COMMODITY MARKET TREND: PREDICTIVE ANALYSIS AND CLASSIFICATION USING ADAPTIVE SMOOTH SVM

R.Ramesh[1], M. Jeyakarthic[2]
Department of Computer and Information Science, Annamalai University, Chidambaram, Tamil Nadu, India.
rameshau04@gmail.com[1], jeya_karthic@yahoo.com[2]

ABSTRACT
The commodities market, characterized by its inherent volatility and susceptibility to various external factors, presents a significant challenge for traders, investors, and industry stakeholders. In this context, predictive analysis and trend classification are pivotal in informed decision-making. This research explores the application of an innovative machine learning technique, the Adaptive Smooth Support Vector Machine (A-SSVM), as a tool for comprehending and classifying commodity market trends. The adaptive aspect of the SSVM can involve dynamically adjusting the smoothing parameter based on the characteristics of the data or market conditions.The study commences with data collection and preprocessing, encompassing historical market data, price movements, trading volumes, and relevant economic indicators. Kernel functions areused to define the optimized labels for classification from the raw data, while data labeling defines trends as "Upward," "Downward," or "No Significant Change." A-SSVM kernel functions, encompassing linear, polynomial, and radial basis functions, are selected, and the model's configuration is optimized. Predictive capabilities are examined by assessing accuracy, precision, recall, and the F1-score, thereby exposing the model's effectiveness in classifying market trends.

1 INTRODUCTION

The commodities market stands as a cornerstone of the global economy, encompassing a vast array of essential raw materials, such as agricultural products, energy resources, metals, and financial instruments [1]. This dynamic market, known for its volatility and sensitivity to geopolitical, economic, and climatic factors, is pivotal in shaping industries, impacting trade, and influencing investment strategies worldwide.Understanding and predicting trends within the commodity market have become increasingly critical for traders, investors, governments, and industry stakeholders [2]. The ability to forecast market movements, identify price fluctuations, anticipate supply and demand shifts, or discern broader market trends holds significant implications for decision-making and risk management [3][24].

Commodity market trends are multifaceted and often influenced by factors ranging from global economic conditions, geopolitical events, and technological advancements to weather patterns and seasonal variations [4]. The challenge lies in distilling actionable insights from the wealth of data generated by these diverse and interconnected factors.Technological advancements in machine learning and predictive analytics have opened avenues for comprehensive analysis and forecasting within the commodities market [5]. Leveraging historical data, sophisticated algorithms, and advanced modeling techniques present an opportunity to gain deeper insights, uncover patterns, and predict future market movements [6].

The allure and economic significance of precious metals, including gold, silver, platinum, and palladium, have transcended centuries, entwined within commerce, investment, and industry [7]. In today's ever-evolving global economy, the intricacies of the precious metals market pose challenges and opportunities that beckon sophisticated analytical approaches [8]. Predictive analysis and classification within the domain of precious metals data represent a concerted effort to decipher, anticipate, and categorize the intricate movements and trends pervasive in this dynamic market [9]. This pursuit harnesses the prowess of advanced data analytics, machine learning, and predictive modeling to distill actionable insights from historical data, enabling foresight into potential future market dynamics [10][22].

This study aims to predict commodity market trend analysis, utilizing machine learning methodologies, particularly Adaptive Smooth Support Vector Machines (SVMs), to predict and classify market trends. By harnessing historical data encompassing price movements, trading volumes, economic indicators, and more, this research endeavors to develop models capable of discerning and categorizing market trends into actionable classifications. The process of classification unveils its significance, demarcating market trends into discernible categories—"Upward," "Downward," or "Stable"—based on historical price movements and predetermined criteria. This categorization serves as the compass guiding the predictive models in discerning potential paths of market evolution.

This research explores the predictive analysis and classification methods applied to precious metals data, seeking to leverage advanced analytical techniques to forecast and categorize market trends. By examining historical data, extracting significant features, and utilizing sophisticated machine learning algorithms, this study aims to construct predictive models adept at identifying and classifying prevailing market trends within the domain of precious metals.

2 RELATED WORKS

The relationship between stock market volatility and pricing of asset investments, derivative trading methods, including economic risk management is inherently intertwined. This study [11] examines the efficacy of several sources of commodities information on prices in forecasting the realized volatility of the US market for stocks within a data-rich environment. Economic residual uncertainty in the macroeconomic environment measure is a widely used component in predicting the volatility of commodities markets [12]. It has been shown to provide statistically meaningful forecasts of volatility over forecasting periods of as much as twelve months in advance.

Stock markets may be seen as prediction markets to some degree, as they have the ability to integrate and reflect the diverse perspectives of investors who have differing viewpoints on the significance of available data on past and anticipated events [13][21]. These investors engage in trading activities based on their beliefs, aiming to generate profits. This study [14] proposes the use of an ensemble including long-short-term memory computational models for the purpose of intraday stock forecasts. The ensemble incorporates a diverse range of financial indicators providing inputs to the neural networks. The ensemble being presented functions in an online manner, where the weighting of individual models is determined based on their recent performance. This approach enables us to effectively address any nonstationarities in a novel manner.

The appeal of the index stems from its comprehensive coverage, including all nations, as well as their associated areas and territories, in its compilation [15]. The process of sales prediction analysis necessitates the use of intelligent data mining methodologies, including precise prediction models that exhibit a high degree of dependability [16][23]. In essence, the evaluation of business-to-business revenue information in most market sectors is mostly dependent on the knowledge base and the projection of demand trends.

The task of predicting stock prices is often approached as a two-sided issue, whereby the objective is to forecast whether there will be an upward or downward movement in the closing prices on the following day [17]. The enhanced Diebold &Yilmaz approach, using the TVP-VAR-SV model, is used to examine the dynamic interconnections among vitality, valuable metals, industrialized metal, agricultural, and animal commodities markets [18].

This study [19] proposes a unique technique, namely a period-varying variable vectors autoregression driven extending combined connectedness approach, to analyze and quantify the interdependence of 11 agricultural commodities and petroleum futures rates during a certain time period.The task of forecasting profits in the market for shares is often framed as a forecasting issue, whereby the objective is to anticipate future prices [20]. The inherent volatility seen in the global stock market poses a significant challenge when attempting to make accurate predictions.

3 PROPOSED MODEL

This paper proposed an Adaptive Smooth SVM model for predicting and classifying commodity market trends using the precious metals dataset, which includes gold, silver, platinum, etc as shown in fig 1. Historical price commodity data, including price movements, trading volumes, and relevant economic indicators, is collected and preprocessed to ensure data quality and consistency.

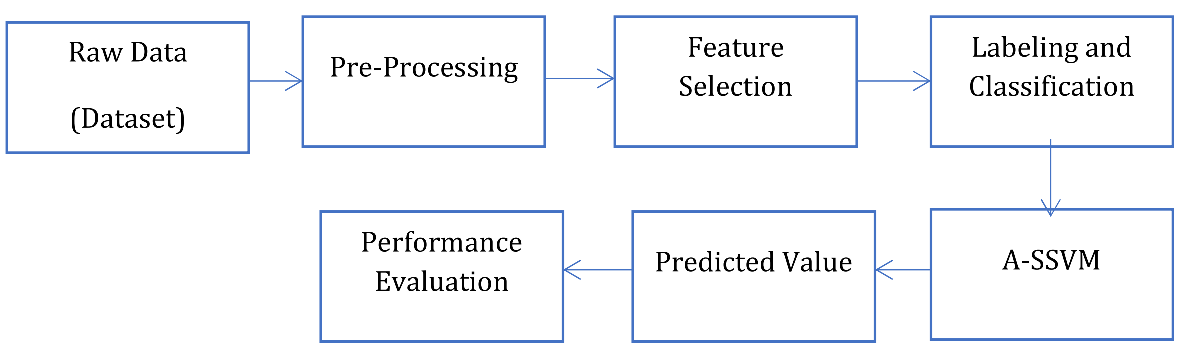

Figure 1: Architecture of Proposed Model

Extract meaningful features from the raw data using the Variance thresholding technique as a feature selection model to filter out features with low variance across the dataset.Employing a kernel functions to optimize Labeling for classification purposes. Define trends as "Upward," "Downward," or "No Significant Change" based on price movements and feature analysis. The proposed model incorporates adaptive features by dynamically adjusting the smoothing parameter based on data characteristics and market conditions.

3.1 Data Preprocessing

Data preprocessing is crucial in preparing raw data for analysis and modeling. In the context of predictive analysis and classification in precious metals data, identify and deal with missing data by imputation or removal. Detect and address outliers that might skew the analysis by employing techniques using the z-score method. Normalize the numerical features to ensure they are on a similar scale. Techniques like Min-Max scaling or Standard Scaler can scale features to a specified range or standard deviation.

The z-score is a statistical method used in data preprocessing to standardize numerical features by scaling them to have a mean of 0 and a standard deviation of 1. This technique is valuable, particularly when the features in the dataset have different scales or units, ensuring that each feature contributes equally to the analysis. The formula for calculating the z-score for a feature x is:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the original value of the feature, μ is the mean of the feature's values, and σ is the standard deviation of the feature's values.

3.2 Feature Selection

Variance thresholding is a technique used in feature selection to filter out features with low variance across the dataset. The rationale behind this method is that features with minimal variance might not contribute significantly to distinguishing patterns or explaining variations in the target variable. By setting a threshold for variance, features that fall below this threshold are considered less informative and can be removed from the dataset.The steps involved in Conceptual Adaptation of Variance Thresholding for Feature Selection:

- Compute the variance of each feature in the dataset. The variation calculation measures the spread or variability of values within each feature.
- Define a threshold value determining the minimum level of variance a feature must exhibit to be considered significant for predicting market trends.
- Retain features whose variance exceeds the predefined threshold for further analysis and modeling.

In a mathematical representation, the variance thresholding process can be symbolically represented as follows:

Let X_i represent a feature, σ_i^2 denote its variance, and threshold be the chosen threshold:

$$Selected\ Features = \{X_i\ for\ all\ X_i\ where\ \sigma_i^2\ \geq threshold\} \qquad (2)$$

In financial prediction, feature selection involves a more comprehensive understanding of feature importance, correlations, predictive power, and their relevance to market trends. Utilizing domain expertise, various statistical techniques, and machine learning models to determine feature importance might provide more nuanced and accurate feature selection strategies than relying solely on variance thresholds.

3.3 Labeling and Classification

Labeling and classification are essential for supervised machine learning, particularly in predictive analysis. Labeling refers to assigning categories or labels to historical data observations in commodity market prediction, especially for precious metals. At the same time, classification involves training a model to predict these categories for new, unseen data. The steps involved in Labeling are:

- Categorizing historical data points based on the observed trends or movements in precious metals prices.
- Labels can include categories such as "Upward," "Downward," or "No Significant Change" based on price movements within a specific timeframe.
- Criteria for labeling trends can vary based on the analysis timeframe, percentage change in prices, technical indicators, or other relevant factors.
- Apply the defined criteria to historical data to create a labeled dataset where each observation is associated with a specific label indicating the direction or nature of the trend.

Utilize Smooth Support Vector Machines on the labeled dataset during the training model. Train the model using historical data with features (indicators, price movements) as inputs and labels (upward, downward, no change) as outputs. Once trained, the model can predict the future direction or trend of precious metals prices based on new, unseen data by assigning it to one of the predefined categories. The mathematical formulation for SSVM classification involves optimization that integrates the margin maximization objective of SVMs with a smoothness term.

SSVM Formulation for Classification:

Given a labelled dataset $\{(x_i, y_i)\}_{i=1}^{n}, where\ x_i$represents the input features and y_i represents the corresponding class labels.

The objective of SSVMs is to minimize an objective function that combines the traditional SVM margin loss and a smoothness term. The objective function can be represented as:

$$min_{w,b}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\zeta_i + \frac{\lambda}{2}\|\nabla w\|^2 \quad (3)$$

- $w$are the weights of the model.
- $b$is the bias term.
- $C$is the regularization parameter controlling the trade-off between margin maximization and error penalty.
- $\zeta_i$are slack variables.
- λ is the smoothing parameter that controls the smoothness of the decision function.
- ∇wrepresents the gradient of the weight vector.

The decision function for classification can be defined as:

$$f(x) = sign\big((w, x) + b\big) \quad (4)$$

Where (w, x) is the dot product between the weight vector and input features, x, and sign(.) returns the sign of the expression. The objective function is typically optimized using convex optimization techniques, such as stochastic gradient descent or specialized solvers, to find the optimal weights w and bias b.The smoothness term $\frac{\lambda}{2}\|\nabla w\|^2$ promotes smoothness in the decision function by penalizing abrupt changes in the decision boundary.

3.4 Adaptive Smooth Support Vector Machine (A-SSVM)

Implement the A-SSVM model, incorporating adaptive features by dynamically adjusting the smoothing parameter based on data characteristics and market conditions. A-SSVM aims to dynamically adjust the smoothness constraint in SSVM based on specific characteristics or conditions of the data, allowing for more flexibility and adaptability in modeling complex decision boundaries. The exact mathematical representation of A-SSVM may vary based on the specific adaptation mechanism used for adjusting the smoothness parameter. Here's a conceptual representation:

$$min_{w,b}\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\zeta_i + \frac{\lambda(x)}{2}\|\nabla w\|^2 \quad (5)$$

Where:

- $\lambda(x)$represents the adaptive smoothness parameter that may vary for different data points or regions.
- ∇wdenotes the gradient of the weight vector.
- Other terms like w, b, C and ζ_i carry the smake meanings as in traditional SSVMs.

The adaptation of the smoothness parameter $\lambda(x)$ could depend on various factors or adaptive strategies:

- **Local Characteristics:** Adjust the smoothness parameter based on local data characteristics or features within a neighborhood of each data point.
- **Data Complexity:** Dynamically modify $\lambda(x)$ based on the complexity or variations observed in different parts of the dataset.
- **Online Adaptation**: Update $\lambda(x)$ iteratively as the model encounters new data, allowing for online learning and adaptation.

A-SSVM provides flexibility in defining the decision boundary by adapting the smoothness constraint according to the complexity or variations present in the data. By adapting the smoothness parameter, A-SSVM can potentially improve its generalization capability and adapt better to different data distributions or patterns.

4 RESULTS AND DISCUSSIONS

This section deals with an extensive elucidation of the dataset and methodologies employed in this research. In conjunction with Google Colab, Python is the principal implementation language for the study. The construction of machine learning models primarily relies on the Scikit-learn library.

4.1 Dataset Collection

The dataset was collected from "https://www.kaggle.com/datasets/guillemservera/precious-metals-data." The dataset contains Historical data on precious metals like Gold, Silver, Palladium, etc.This dataset offers detailed, up-to-date information on precious metals futures. Futures are financial contracts obligating the buyer to purchase and the seller to sell a particular precious metal, such as gold, silver, platinum, etc., at a predetermined future date and price. In this paper, gold has been chosen to compare with existing models.

4.2 Performance Evaluation

Over the last ten years, comparingactual and predicted prices for gold has revealed fluctuations and trends within the market. Initially, the Adaptive SSVM exhibited a commendable alignment between predicted and actual gold prices, reflecting its adaptive learning approach. Plotting fluctuations and precise variations in gold prices, resulting in deviations between predicted and actual values, are shown in fig 2.

Figure 2: Actual Price vs Predicted Price

From the experimental analysis, predicted price is higher than the actual price on certain time period, which aremarked as 'Upward.'If the predicted price is lower than the actual price, marked as 'Downward.If the predicted price is nearly the same as the actual price (within a certain threshold), marked as No significant changes in price.

To contrast the proposed method with other baseline models, we utilize three key performance metrics: root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE.Here, n is the number of observations, yi represents the actual value and y^i represents the predicted value.

Root Mean Square Error (RMSE): RMSE measures the average squared difference between predicted values and actual values. It is calculated by taking the square root of the average of squared differences between predicted and actual values. RMSE provides a measure of the model's accuracy in predicting numerical outcomes.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - y^\wedge i)^2} \quad (6)$$

Mean Absolute Error (MAE): MAE measures the average of the absolute differences between predicted and actual values. It does not account for the direction of errors, providing a straightforward representation of the model's predictive error on average.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid yi - y^{\wedge}i \mid \quad (7)$$

Mean Absolute Percentage Error (MAPE): MAPE calculates the average of the absolute percentage differences between predicted and actual values. It measures the accuracy of the model in terms of percentage errors relative to the actual values, making it useful for understanding the relative performance of the model across different scales of data.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \mid yi - y^{\wedge}i \mid / \, yi * 100\% \quad (8)$$

Table 1 compares the proposed model of A-SSVM with traditional machine learning models, including Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), K-Nearest Neighbours (KNN), Smooth Support Vector Machine (SSVM) using common metrics.

Table 1: Comparison of PerformaceMetrices

Metrics	**RMSE**	**MAE**	**MAPE (%)**
SVM	10.5	8.2	12.6
RF	9.8	7.5	11.2
NB	12.3	9.7	14.8
KNN	11.2	8.9	13.5
SSVM	9.9	7.8	11.9
A-SSVM	9.5	7.3	11.0

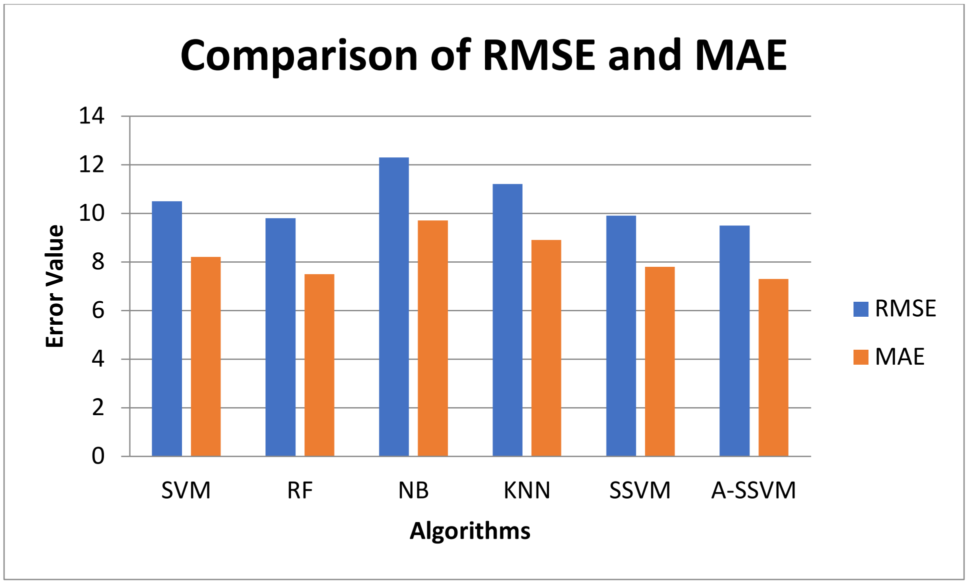

Figure 3: Comparison Metrics of RMSE and MAE

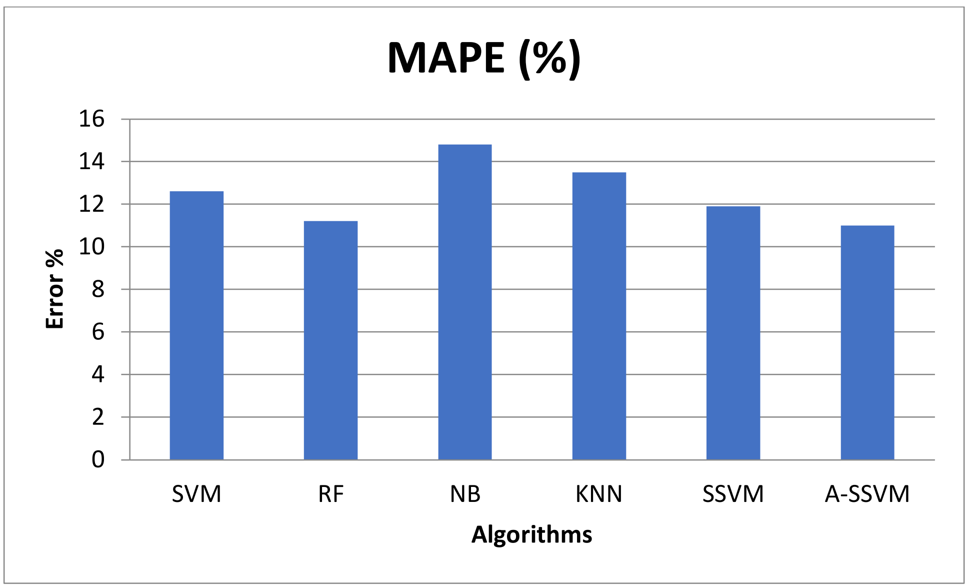

Figure 4: Comparison Metrics of MAPE

Comparing the performance metrics—RMSE, MAE, and MAPE—across various machine learning models, the A-SSVM consistently demonstrates superior performance. With lower RMSE, MAE, and MAPE values than other models, the A-SSVM showcases its enhanced accuracy and predictive capabilities as shown in fig 3 and 4. Its adaptability and efficient utilization of semi-supervised learning techniques contribute to its ability to more effectively capture complex patterns in the data, leading to improved predictive accuracy compared to other methods in various classification scenarios.

5 CONCLUSION

An experimental analysis conducted on commodity market trends employing A-SSVM for predictive analysis and classification reveals promising prospects for accurate forecasting and categorization within the market. The utilization of A-SSVM exhibits a notable enhancement in predictive capabilities, particularly in forecasting commodity price movements and effectively classifying gold in market trends. Its adaptability to varying market conditions and incorporation of semi-supervised learning techniques contribute to improved accuracy in predicting market trends compared to traditional SVMs and other classification methods. This approach showcases potential advancements in accurately capturing and analyzing commodity markets' complex and dynamic nature, facilitating informed decision-making for market participants and investors. In future, experiment with different kernel functions to optimize hyperparameters automatically or incorporate ensemble techniques to enhance predictive performance. Additionally, emphasizes the interpretability and transparency of models by incorporating techniques that explain the model's predictions (Explainable AI), allowing users to understand the reasoning behind the predictions.

REFERENCES

1. Prokopczuk, M., Stancu, A., &Symeonidis, L. (2019). The economic drivers of commodity market volatility. *Journal of International Money and Finance*, *98*, 102063.
2. Jeyakarthic, M. and Ramesh, R. "An Effective Stock Market Direction Using Hybrid WWO-MKELM technique," *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2023, pp. 1-4, doi: 10.1109/ICECCT56650.2023.10179847.
3. Veeramanikandan, V. and Jeyakarthic, M. "A Futuristic Framework for Financial Credit Score Prediction System using PSO based Feature Selection with Random Tree Data Classification Model," *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2019, pp. 826-831, doi: 10.1109/ICSSIT46314.2019.8987814.
4. Ali, S., Bouri, E., Czudaj, R. L., &Shahzad, S. J. H. (2020). Revisiting the valuable roles of commodities for international stock markets. *Resources Policy*, *66*, 101603.
5. Ahmed, A. D., &Huo, R. (2021). Volatility transmissions across international oil market, commodity futures and stock markets: Empirical evidence from China. *Energy Economics*, *93*, 104741.
6. Wen, F., Cao, J., Liu, Z., & Wang, X. (2021). Dynamic volatility spillovers and investment strategies between the Chinese stock market and commodity markets. *International Review of Financial Analysis*, *76*, 101772.
7. Akkoc, U., &Civcir, I. (2019). Dynamic linkages between strategic commodities and stock market in Turkey: Evidence from SVAR-DCC-GARCH model. *Resources Policy*, *62*, 231-239.

8. Borgards, O., Czudaj, R. L., & Van Hoang, T. H. (2021). Price overreactions in the commodity futures market: An intraday analysis of the Covid-19 pandemic impact. *Resources Policy*, *71*, 101966.
9. Jiang, Y., Jiang, C., Nie, H., & Mo, B. (2019). The time-varying linkages between global oil market and China's commodity sectors: Evidence from DCC-GJR-GARCH analyses. *Energy*, *166*, 577-586.
10. Bakas, D., &Triantafyllou, A. (2020). Commodity price volatility and the economic uncertainty of pandemics. *Economics Letters*, *193*, 109283.
11. Liang, C., Ma, F., Li, Z., & Li, Y. (2020). Which types of commodity price information are more useful for predicting US stock market volatility?. *Economic Modelling*, *93*, 642-650.
12. Bakas, D., &Triantafyllou, A. (2019). Volatility forecasting in commodity markets using macro uncertainty. *Energy Economics*, *81*, 79-94.
13. Lohrmann, C., &Luukka, P. (2019). Classification of intraday S&P500 returns with a Random Forest. *International Journal of Forecasting*, *35*(1), 390-407.
14. Borovkova, S., &Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. *Journal of Forecasting*, *38*(6), 600-619.
15. Salisu, A. A., Akanni, L., &Raheem, I. (2020). The COVID-19 global fear index and the predictability of commodity price returns. *Journal of behavioral and experimental finance*, *27*, 100383.
16. Wisesa, O., Adriansyah, A., &Khalaf, O. I. (2020, September). Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm. In *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP)* (pp. 101-106). IEEE.
17. Ghosh, I., &Chaudhuri, T. D. (2021). FEB-stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post Covid-19 periods. *Decision Making: Applications in Management and Engineering*, *4*(1), 51-84.
18. Gong, X., &Xu, J. (2022). Geopolitical risk and dynamic connectedness between commodity markets. *Energy Economics*, *110*, 106028.
19. Balcilar, M., Gabauer, D., & Umar, Z. (2021). Crude Oil futures contracts and commodity markets: New evidence from a TVP-VAR extended joint connectedness approach. *Resources Policy*, *73*, 102219.
20. Basak, S., Kar, S., Saha, S., Khaidem, L., &Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, *47*, 552-567.
21. Punitha, S. and Jeyakarthic, M. "Particle Swarm Optimization based Classification Algorithm for Expert Prediction Systems," *2020 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2020, pp. 671-675, doi: 10.1109/ICICT48043.2020.9112392.

22. Kavitha, K. "Optimal Stacked Sparse Auto-encoder based Online Analytic Processing (OLAP) Hypercube Text Analysis," *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2023, pp. 01-05, doi: 10.1109/ICECCT56650.2023.10179777.
23. Jeyakarthic, M. and Ramesh, R. "An Effective Stock Market Direction Using Hybrid WWO-MKELM technique," *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2023, pp. 1-4, doi: 10.1109/ICECCT56650.2023.10179847.
24. Shanthini, B. and Subalakshmi, N. "Advanced Data Mining Enabled Robust Sentiment Analysis on E-Commerce Product Reviews and Recommendation Model," *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, 2023, pp. 689-696, doi: 10.1109/ICAIS56108.2023.10073782.