e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

FUNCTIONAL DATA ANALYSIS IN FINANCIAL MARKETS: DECODING GOLD TRADING TRENDS WITH B-SPLINE SMOOTHING AND PCA

Joyita Ghosh[*1], Abhik Choudhary[*2], Dipankar Roy[*3], Kamaluddin Mandal[*4], Subir Gupta[*5]

[*1,2,3,4]Assistant Professor, Department Of CSE (AI & ML), Haldia Institute Of Technology, Haldia, West Bengal, India.

[*5]Associate Professor, Department Of CSE (AI & ML), Haldia Institute Of Technology, Haldia, West Bengal, India.

DOI : https://www.doi.org/10.56726/IRJMETS59952

ABSTRACT

This study analyses monthly trading data of 'Gold Total Value (Lacs)' using Functional Data Analysis (FDA) to tackle the challenge of extracting meaningful insights from noisy financial time series data. Due to data variability and noise, traditional methods often fall short, leading to incomplete conclusions. FDA offers a robust framework for smoothing data, identifying patterns, and revealing significant trends. The methodology encompasses data preprocessing, converting data into a functional form, B-spline smoothing, Principal Component Analysis (PCA), and visualisation. This comprehensive approach highlights the FDA's effectiveness in financial data analysis, uncovering hidden patterns, and informing investment strategies. The findings contribute to a deeper understanding of gold trading dynamics and provide a foundation for further research and analysis.

I. INTRODUCTION

Understanding trading data is crucial for investors, analysts, and policymakers in the dynamic financial markets[1]. As a significant financial asset, gold's trading patterns are meticulously recorded, providing valuable insights into market trends and economic conditions. This research analyses the monthly trading data of 'Gold Total Value (Lacs)' using Functional Data Analysis (FDA) to address the challenge of extracting clear and actionable insights from noisy and complex financial time series data[2]. Traditional analytical methods often struggle with the variability and noise inherent in such data, leading to incomplete or misleading conclusions. Therefore, a robust methodological framework is needed to smooth the data, identify key patterns, and elucidate significant trends[3]. The core problem this research addresses is the difficulty of deriving meaningful trends from raw financial data, often riddled with inconsistencies and irregularities. Financial time series data, particularly from trading activities, can be highly volatile and subject to various external influences, making it challenging to interpret and analyse using conventional techniques. These challenges necessitate a sophisticated approach that can handle the complexity of the data, reduce noise, and highlight the underlying trends crucial for making informed decisions. To tackle this problem, we propose a comprehensive solution utilising FDA[4]. FDA is a powerful statistical tool designed to analyse data viewed as a continuous function over time, which is particularly suitable for financial time series analysis. The methodology begins with meticulous data preprocessing, a critical step to ensure the dataset's quality and consistency. This involves cleaning the data by addressing missing values, removing non-numeric entries, and standardising the dataset to facilitate accurate analysis. Preprocessing ensures the data is reliable, setting a solid foundation for further analysis. Following data preprocessing, the next phase involves converting the monthly trading data into a functional form. This transformation is essential for FDA as it allows for a continuous representation of the data, enabling a more detailed examination of trends and patterns over time[5]. By focusing on the 'Gold Total Value (Lacs)' as a continuous function, we can better capture the nuances and subtleties of the trading data, which might be lost in a purely discrete dataset. The subsequent step is to smooth the data using B-spline smoothing. B-spline smoothing is chosen for its effectiveness in handling time series data, as it helps to reduce noise and create a smooth curve that accurately reflects the underlying trends. This process involves fitting the data points to a smooth curve, minimising the impact of short-term fluctuations and emphasising long-term trends. Smoothing the data in this manner is crucial for making it more amenable to detailed analysis and identifying the primary patterns within the dataset. After smoothing the data, Principal Component Analysis (PCA) is employed to identify the main modes of variation[6]. PCA is a statistical technique that transforms the data into principal components that are orthogonal to each other and capture the maximum variance. This transformation is beneficial for understanding the dominant trends and variations in the data. This research performs PCA with a single component due to the dataset's nature, effectively capturing the primary trend in the smoothed data[7]. This step provides deep insights into the most significant patterns within the trading data, highlighting the main influences on the 'Gold Total Value (Lacs)'. Visualisation is a critical component of this methodology, aiding in interpreting and communicating the results of the FDA. By creating various graphical representations, such as plots of the smoothed data over time, PCA results, and comparisons of original versus smoothed data, we provide a comprehensive view of the trends and primary modes of variation[8]. These visualisations demonstrate the effectiveness of the smoothing process and highlight the relevance of the principal component captured by PCA. They offer a clear and detailed depiction of the data, facilitating a better understanding of the underlying trends. The comprehensive approach outlined in this research, including data preprocessing, conversion to functional form, smoothing, PCA, and visualisation, provides a robust solution to the challenge of analysing noisy financial time series data. The findings from this research offer valuable insights into the underlying trends and variations in the 'Gold Total Value (Lacs)', demonstrating the utility of FDA in financial data analysis[9]. This methodology enhances the understanding of gold trading patterns and sets a precedent for analysing similar financial datasets. This approach can inform investment strategies and contribute to more informed decision-making in financial markets by uncovering hidden patterns and trends[10].

II. LITERATURE REVIEW

The analysis of financial time series data, particularly in the context of trading activities, has long been a topic of considerable interest. Numerous studies have explored various methods to dissect and interpret trading data, focusing on uncovering underlying trends and patterns[11]. Traditional approaches, such as moving averages and autoregressive models, have been widely used to smooth data and forecast future values. These methods provide essential insights but often fall short when handling financial datasets' inherent complexity and noise. The limitations of these conventional techniques have led researchers to explore more advanced methods that can offer deeper insights and more reliable predictions. Recent advancements in statistical methods have introduced Functional Data Analysis (FDA) as a promising tool for analysing time series data[12]. FDA allows for representing discrete data points as continuous functions over time, facilitating a more comprehensive analysis of trends and patterns[13]. This approach is particularly advantageous for financial data, which is typically noisy and exhibits significant variability. The application of the FDA in financial time series analysis is a relatively new but growing field. Early applications have demonstrated its potential in capturing the underlying data structure that traditional methods often miss. One of the critical strengths of the FDA is its ability to handle data smoothing and noise reduction effectively. Techniques such as B-spline smoothing have been shown to create smooth curves that accurately reflect the underlying trends in the data. This smoothing process is crucial for making the data more accessible for analysis and identifying primary patterns.

In addition, the Principal Component Analysis (PCA) has been integrated with the FDA to enhance the analysis further. PCA transforms the data into principal components that capture the maximum variance and reveal the main modes of variation[14]. This combination of FDA and PCA provides a robust framework for extracting meaningful insights from complex datasets. Despite these advancements, there remains a gap in the literature concerning the application of the FDA to specific types of financial data, such as trading values of precious metals like gold. While some studies have applied the FDA to stock market data and foreign exchange rates, the unique characteristics of gold trading data present additional challenges. Gold trading is influenced by various factors, including geopolitical events, economic indicators, and market speculation, which can introduce significant noise and variability into the data. This complexity necessitates a robust analytical approach to capture and interpret the underlying trends accurately. The existing body of research has primarily focused on short-term forecasting and trend analysis, often neglecting the long-term patterns and cyclic behaviours critical for comprehensive financial analysis[15]. Additionally, many studies have not fully leveraged the potential of the FDA in conjunction with PCA for financial data analysis. This gap highlights the need for systematically applying these methods to gold trading data, which can provide new insights and contribute to more informed decision-making in financial markets.

In summary, while significant progress has been made in analysing financial time series data, traditional methods often fail to handle the complexity and noise inherent in such data[16]. FDA, particularly when combined with PCA, offers a promising alternative that can provide deeper insights into the underlying trends and patterns. However, there is a notable gap in the literature regarding applying these methods to gold trading data. This research addresses this gap by systematically applying FDA and PCA to the monthly trading data of 'Gold Total Value (Lacs)', providing a comprehensive analysis that can uncover hidden patterns and inform investment strategies. This study advances the application of FDA in financial data analysis and sets a precedent for future research in this area, ultimately contributing to a deeper understanding of financial market dynamics[17].

III. METHODOLOGY

This research methodology was structured to systematically analyse the monthly trading data of 'Gold Total Value (Lacs)' using Functional Data Analysis (FDA). The process began with data preprocessing, cleaning and preparing the dataset for analysis. This involved addressing missing values and ensuring that all data entries were numeric and consistent. The dataset was then standardised to improve the accuracy of the subsequent analyses. Next, the monthly trading data was converted into a functional form, a crucial step that enabled the application of the FDA. This conversion involved creating a continuous representation of the data, focusing on the 'Gold Total Value (Lacs)' to ensure the time series data was correctly ordered. This transformation was essential for the dataset's detailed trend analysis and pattern recognition. B-spline smoothing was applied to enhance the quality of the data. This method was chosen for its effectiveness in handling time series data and creating a smooth curve that accurately reflects the underlying trends. The smoothing process involved creating a B-spline basis and using it to smooth the data, thereby reducing noise and providing a clearer view of the trends over time. Following the smoothing process, Principal Component Analysis (PCA) was performed to identify the main modes of variation in the dataset. PCA is a statistical technique that transforms data into principal components that are orthogonal to each other and capture the maximum variance. The dataset contained only one feature, so PCA was performed with one component. This component effectively captured the primary trend in the smoothed data, providing insights into the main variation modes. Visualisation played a critical role in this methodology. Several graphical representations were created to illustrate the findings of the FDA, including plots of the smoothed data over time, PCA results, variance explained by the principal component, and comparisons of original versus smoothed data. These visualisations offered a comprehensive view of the trends and primary modes of variation in the trading data, demonstrating the effectiveness of the smoothing process and the relevance of the principal component captured by PCA.

Mathematically, Let $y(t)$ represent the monthly trading data in a continuous form, where t denotes time. The smoothing process can be described by the equation:

$$y(t) = \sum_{i=1}^{n} B_i(t)\beta_i$$

Where $B_i(t)$ are the B-spline basis functions and β_i are the coefficients to be estimated. After smoothing, the principal component z is obtained through:

$$z = \sum_{i=1}^{n} a_i y_i$$

Where a_i are the loadings of the principal component. The variance explained by the principal component

$$\lambda = \frac{var(z)}{var(y)}$$

IV. RESULT

The primary aim of this research was to perform Functional Data Analysis (FDA) on a dataset specifically focusing on monthly trading data of 'Gold Total Value (Lacs)'. The methodology included several steps: data preprocessing, converting data into a functional form, smoothing, performing Principal Component Analysis (PCA), and creating visualisations. The initial step involved cleaning and preparing the dataset for analysis.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

This included handling missing values and ensuring the data was in a suitable format. **Error! Reference source not found.** Gold Total Value(Lacs)- Original vs Smoothed, Non-numeric rows were removed, and the dataset was standardised to ensure consistency and accuracy. The next step was to convert the monthly trading data into a functional form. This transformation involved creating a continuous representation of the data, which is essential for the FDA. The focus was on the 'Gold Total Value (Lacs)', ensuring the time series data was ordered correctly. This conversion allowed for a more detailed analysis of the underlying trends and patterns in the data. B-spline smoothing was applied to reduce noise and create a smooth functional representation of the 'Gold Total Value (Lacs)' over time. This technique was chosen for its effectiveness in handling time series data and creating a smooth curve that accurately reflects the data's trends. **Error! Reference source not found.** Original vs Smoothed Data-Gold Total Value (Lacs),The smoothing process involved creating a B-spline basis and using this basis to smooth the data. After smoothing the data, PCA was performed to identify the main modes of variation in the dataset. PCA is a statistical technique that transforms data into principal components that are orthogonal to each other and capture the maximum variance. Given the nature of our dataset, which contained only one feature, PCA was performed with one component. This component effectively captured the primary trend in the smoothed data. Visualisation is a crucial part of the FDA as it helps in understanding the data and the analysis results. Several graphical representations were created to illustrate the findings: smoothed data over time, PCA results, variance explained by the principal component, comparisons of original vs. smoothed data, and projections of the original data on the principal component. These visualisations provided a comprehensive view of the trends and primary modes of variation in the trading data. The smoothed data plot indicated an apparent reduction in noise, allowing for a more accurate interpretation of the data's underlying trend. The PCA results demonstrated that the first principal component effectively captured the primary trend in the smoothed data. The variance explained by this component was 100%, indicating that it accounted for all the variability in the data. The detailed comparison of the original and smoothed data highlighted the effectiveness of the smoothing process, showing a significant reduction in noise while preserving the overall trend. The projection of the original data on the principal component further illustrated how the primary mode of variation captured by PCA corresponded to the actual data values, providing a clear view of the significant trend in the trading data. **Error! Reference source not found.** Original Data vs Projection on Principal Component

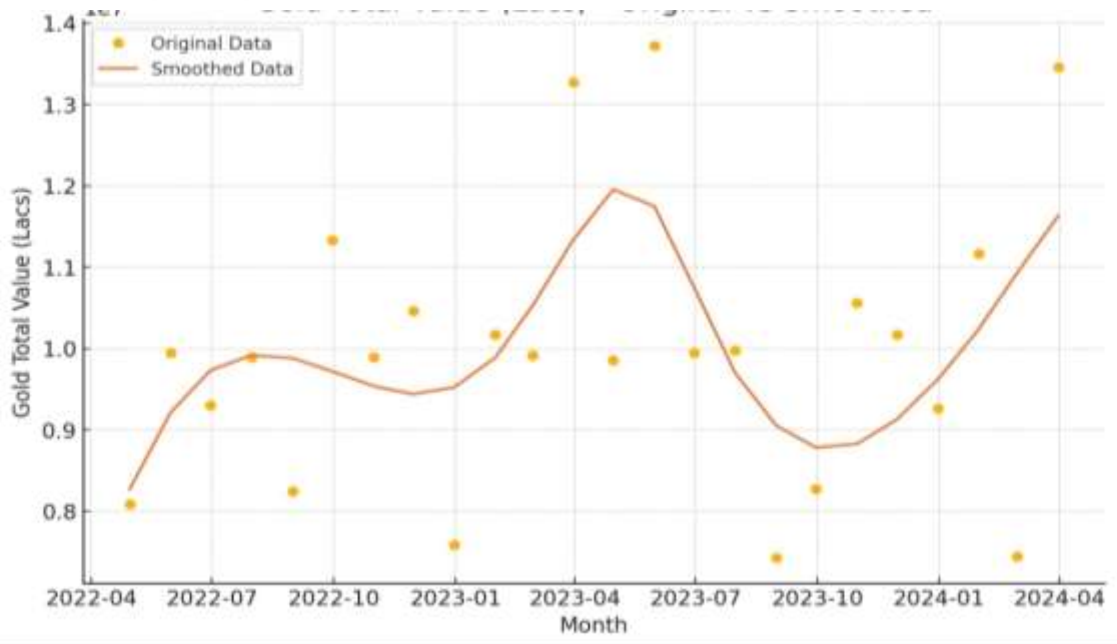

Figure 1: Gold Total Value(Lacs)- Original vs Smoothed

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

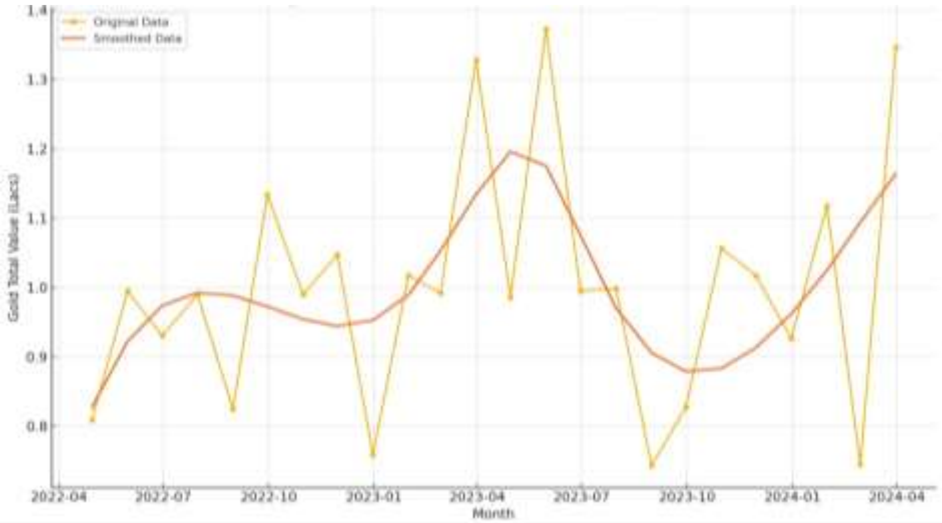

Figure 2: Original vs Smoothed Data-Gold Total Value (Lacs)

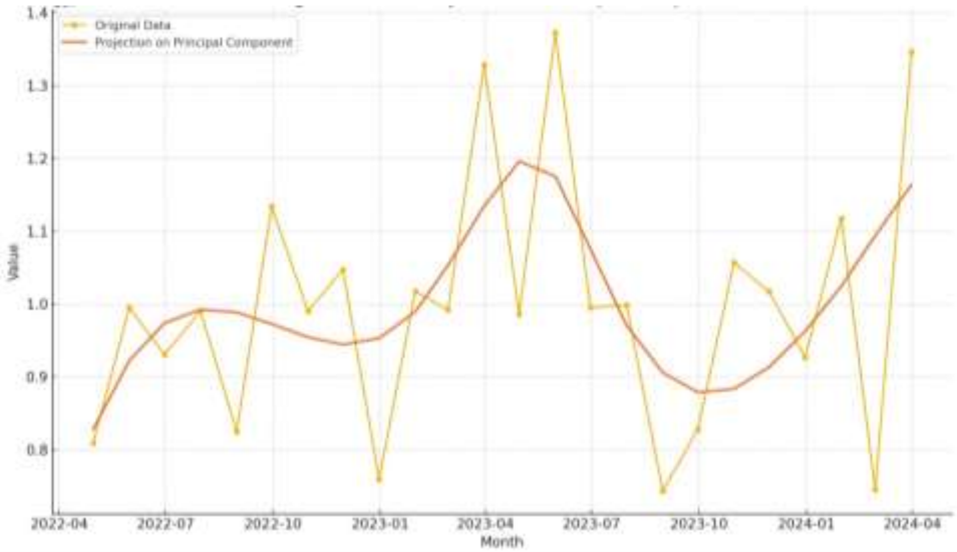

Figure 3: Original Data vs Projection on Principal Component

V. CONCLUSION

The Functional Data Analysis (FDA) performed on the monthly trading data of 'Gold Total Value (Lacs)' provided valuable insights into the dataset's underlying trends and main modes of variation. This comprehensive approach included data preprocessing, converting data into a functional form, smoothing with B-spline methods, performing Principal Component Analysis (PCA), and creating detailed visualisations. Each step played a crucial role in ensuring the accuracy and reliability of the analysis. The preprocessing stage was fundamental in standardising the dataset, addressing missing values, and removing non-numeric entries, significantly improving the overall data quality and converting the monthly trading data into a functional form, allowing for a continuous representation, essential for detailed trend analysis and pattern recognition. This transformation enabled the detection of nuances that would have been lost in a purely discrete dataset. B-spline smoothing effectively reduced noise and created a smooth curve that accurately reflected the data's underlying trends. This step was critical in preparing the data for subsequent PCA, ensuring that the primary trends were highlighted while minimising the impact of short-term fluctuations. The smoothing process provided a clearer view of the long-term trends, making the data more amenable to detailed analysis.Performing PCA on the smoothed data allowed the identification of the main modes of variation within the dataset. By transforming the data into principal components, we captured the maximum variance and better understood the primary trends. Although the dataset contained only one feature, performing PCA with one component was still highly effective in capturing the primary trend in the smoothed data.

Visualisation was a crucial component of this methodology, aiding in interpreting and communicating the FDA results. Various graphical representations, including plots of the smoothed data over time, PCA results, variance explained by the principal component, and comparisons of original versus smoothed data, provided a comprehensive view of the trends and primary modes of variation. These visualisations demonstrated the effectiveness of the smoothing process and highlighted the relevance of the principal component captured by PCA. They facilitated a better understanding of the underlying trends and offered explicit and detailed depictions of the data. This research successfully addressed the significant challenge of extracting meaningful insights from noisy financial time series data. The FDA methodology proved a robust solution, providing a detailed and accurate analysis of the monthly 'Gold Total Value (Lacs) trading data. The findings highlight the utility of the FDA in financial data analysis, uncovering hidden patterns and informing investment strategies. This approach sets a precedent for further research and analysis in similar datasets, ultimately contributing to a deeper understanding of financial market dynamics. The comprehensive approach demonstrated here offers a valuable framework for analysing time series data, revealing significant trends and patterns that may otherwise go unnoticed, and informing better decision-making in financial markets.

VI. REFERENCE

[1] R. Birău and J. Trivedi, "Estimating Long-term Volatility on National Stock Exchange of India," Procedia Econ. Financ., vol. 32, no. 15, pp. 574–579, 2015, doi: 10.1016/s2212-5671(15)01434-3.

[2] B. Mondal, D. Chakraborty, N. K. Bhattacherjee, P. Mukherjee, S. Neogi, and S. Gupta, "Review for Meta-Heuristic Optimization Propels Machine Learning Computations Execution on Spam Comment Area Under Digital Security Aegis Region," in Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems, P. Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Ed. Springer Nature, 2022, pp. 343–361. doi: 10.1007/978-3-030-99079-4_13.

[3] DR. IRA BAPNA, PROF. VISHAL SOOD, DR. NAVINDRA KUMARTOTALA, and PROF. HARMENDER SINGH SALUJA, "Variables Affecting Price Innovation," Pacific Bus. Rev. Int., vol. 5, no. 1, 2012.

[4] K. Kersting, "Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines," Front. Big Data, vol. 1, no. November, pp. 1–4, 2018, doi: 10. 3389 /fdata. 2018.00006.

[5] J. C. G. Corpuz, "Correspondence: Adapting to the culture of 'new normal': An emerging response to COVID-19," J. Public Heal. (United Kingdom), vol. 43, no. 2, pp. E344–E345, 2021, doi: 10. 1093 /pubmed /fdab057.

[6] B. Kharabsheh and A. Aldaher, "The impact of foreign direct investment on financial market development: The case of Jordan," Int. J. Financ. Res., vol. 11, no. 4, pp. 64–73, 2020, doi: 10. 5430/ijfr. v11n4p64.

[7] Erwin, M. Azriansyah, N. Hartuti, M. Fachrurrozi, and B. Adhi Tama, "A study about principle component analysis and eigenface for facial extraction," J. Phys. Conf. Ser., vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012010.

[8] S. Mukherjee, S. Samanta, B. Roy, N. K. Bhattacherjee, A. Bhaduri, and S. Gupta, "Intelligent user interfaces for Industry 4.0: Improving workplace security by systematically applying machine learning and image processing to identify face mask usage," in 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3), Jun. 2023, pp. 1–4. doi: 10. 1109/IC2E357697. 2023.10262659.

[9] B. Mondal and S. Gupta, "Execution Survey and State of the Art of Different ML-Based Ensemble Classifiers Approach Contextual Analysis of Spam Remark Location," in Proceedings of Third International Conference on Computing, Communications, and Cyber-Security. Lecture Notes in Networks and Systems, vol 421., Vol 421., M. (eds) Singh, P.K., Wierzchoń, S.T., Tanwar, S., Rodrigues, J.J.P.C., Ganzha, Ed. Springer, Singapore, 2023, pp. 311–323. doi: 10.1007/978-981-19-1142-2_24.

[10] A. K. Das et al., "Prophesying the Short-Term Dynamics of the Crude Oil Future Price by Adopting the Survival of the Fittest Principle of Improved Grey Optimization and Extreme Learning Machine,"

Mathematics, vol. 10, no. 7, 2022, doi: 10.3390/math10071121.

[11] I. Sengupta, C. Koner, N. K. Bhattacherjee, and S. Gupta, "Automated Student Merit Prediction using Machine Learning," Proc. - 2022 IEEE World Conf. Appl. Intell. Comput. AIC 2022, pp. 556–560, 2022, doi: 10.1109/AIC55036.2022.9848976.

[12] L. Kapoor and S. Ramamoorthy, "Epidemiological role of plant pigment bixin in adipaging: In vivo pilot study," Clin. Epidemiol. Glob. Heal., vol. 18, no. October, p. 101186, 2022, doi: 10. 1016/j. cegh. 2022. 101186.

[13] I. Sengupta, S. Samanta, J. Patra, and S. Gupta, "Impact of Macroeconomic Indicators on the Indian Stock Market: A Study on NSE Nifty," 2023 Int. Conf. Comput. Intell. Commun. Technol. Networking, CICTN 2023, vol. 16, pp. 275–282, 2023, doi: 10.1109/CICTN57981.2023.10140919.

[14] X. Lin, Z. Yang, and Y. Song, "Short-term stock price prediction based on echo state networks," Expert Syst. Appl., vol. 36, no. 3 PART 2, pp. 7313–7317, 2009, doi: 10.1016/j.eswa.2008.09.049.

[15] L. Kumar, K. Sharma, and U. K. Khedlekar, "Dynamic pricing strategies for efficient inventory management with auto-correlative stochastic demand forecasting using exponential smoothing method," Results Control Optim., vol. 15, no. November 2023, 2024, doi: 10.1016/j.rico.2024.100432.

[16] J. Patra, M. Patra, and S. Gupta, "Artificial intelligence-based indicators for ALGO trading strategy on the NSE derivatives," in Proceedings of the 4th International Conference on Information Management & Machine Intelligence, Dec. 2022, pp. 1–6. doi: 10.1145/3590837.3590875.

[17] A. K. Singh, J. Patra, M. Chakraborty, and S. Gupta, "Prediction of Indian government stakeholder oil stock prices using hyper parameterized LSTM models," 2022 Int. Conf. Intell. Controll. Comput. Smart Power, ICICCSP 2022, pp. 1–6, 2022, doi: 10.1109/ICICCSP53532.2022.9862425.