IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

AI-Powered Decision Support System for Coffee Commodity Investment

[1]KOTIPALLI APPANNA,
Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]CH KODANDA RAMU, Miracle Educational Society Group of Institutions
[3]K.SUKESH KUMAR PATNAIK, Miracle Educational Society Group of Institutions
[1]appannakotipalli1234@gmail.com

ABSTRACT:

This work proposes an AI-based decision support system for coffee commodity trading using big data analytics and machine learning techniques. Forecasting and accurately predicting commodity trading is a problem one encounters with traditional trading models. Because of these issues, the proposed system with its ETL pipeline powered by technologies like Hadoop and Spark processes data in real-time and analyzes it. For accurate forecasting of coffee prices and trading activities, advanced ML/DL models like XGBoost and ensemble voting models are also incorporated. Interactive dashboards for visualization of the market and the relationships of the buyers and sellers is another highlighted feature of the system. Through the use of intelligent automation and predictive analytics, the framework aids in risk mitigation, profit optimization, and well-informed decision making by the investors and traders in the ever-changing coffee commodity market.

Keywords: Hadoop and Spark, coffee commodity, XGBoost

INTRODUCTION

Coffee serves as a daily ritual for millions of people globally and it is one of the most traded commodities across the globe. Following oil, coffee holds a significant position in the economies of many countries, especially in Latin America, Africa, and Asia. In other scenarios, the coffee trading market stabilizes during periods of calm. However, the market also experiences some extreme volatility during periods of unstable weather, unpredictable political scenarios, global demand changes, and currency fluctuations. For investors and traders aiming for consistent returns, this volatility creates profound difficulties. Here, Artificial Intelligence (AI) emerges as a transformative force because it can process enormous amounts of data and provide insights well beyond what conventional analytical frameworks can offer. With the aid of big data, as well as machine and deep learning algorithms, traders are equipped with better insights into market dynamics, enabling them to forecast prices and make calculated decisions with lower exposure to risk. This initiative seeks to transform coffee commodity trading through the application of AI.

RELATED WORK

In the last few years, the application of AI and big data into the financial markets, especially in commodity trading, has attracted the attention of many researchers. Le et al. (2024) designed an AI-powered data-driven framework to assist coffee commodity trading. The model focused on guiding decisions through big data, although the challenge of real-time application scalability remained unresolved. Seo et al. (2023) developed an AI-based financial market decision-making system with deep learning capabilities designed for market forecasting. Despite the model achieving a measure of accuracy, its extensive computational requirements rendered it infeasible for many small traders. Zhang et al. (2021) focused on the application of neural networks to stock price forecasting. Their model had precise accuracy in detecting price movements, but it had issues with overfitting and being too sensitive to noisy input data. Kim et al. (2020) used deep learning models to analyze market trends and were able to identify complex trading patterns. However, the model's reliance on large labeled datasets made it difficult to implement in real-time volatile markets such as coffee trading. Patel et al. (2018) proposed a hybrid AI model with multiple algorithms to improve risk management in banking. While the ensemble approach made it more reliable, the model controlling the risk management system's architecture and infrastructure had to be more complex, bringing with it a demand for rigorous model tuning. These previous works have provided the foundation for using AI in the financial markets, but faced challenges in areas such as processing data in real-time, predicting with confidence, and easy deployment. Most did not have strong ETL pipelines with integrated visualization capabilities, crucial for commodities such as coffee where market conditions change rapidly.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	**Contribution**	**Impact on Current Research**
Le et al. (2024)	Applied AI and big data to optimize coffee trading decisions	Highlighted the need for scalable, real-time analytics in coffee commodity markets
Seo et al. (2023)	Developed deep learning models for financial decision-making	Demonstrated predictive power of AI but revealed limitations in accessibility for small traders
Zhang et al. (2021)	Used neural networks for stock price prediction	Inspired use of deep learning in commodity price forecasting while noting overfitting issues
Kim et al. (2020)	Leveraged deep learning for market trend	Validated AI's capability in uncovering

	detection	patterns, encouraging use of complex models
Patel et al. (2018)	Introduced hybrid AI models for banking risk management	Supported ensemble modeling in finance but noted high implementation complexity

PROPOSED APPROACH

This system proposes an integrated AI framework for coffee commodities trading, which improves decision making by combining advanced machine learning (ML) and deep learning (DL) models on big data. The approach follows through the three-layer ETL (Extract, Transform, Load) pipeline to collect, clean, and integrate data from different sources like financial websites, global commodity indexes, and even government reports. Structured and unstructured data are efficiently processed by the system, which runs on Apache Hadoop and Spark, on a global scale. Other important highlights include the use of XGBoost, an optimized gradient boosting algorithm, and the ensemble voting in which multiple ML/DL models' predictions are combined to improve dependability and precision. Dashboards are added to display data concerning buyer-seller interactions, market movements, and forecasted coffee prices, which facilitates understanding of complex datasets through intuitive analytics for traders. Real-time predictions are incorporated into the architecture which assists the investor in making good trades and market risk management.

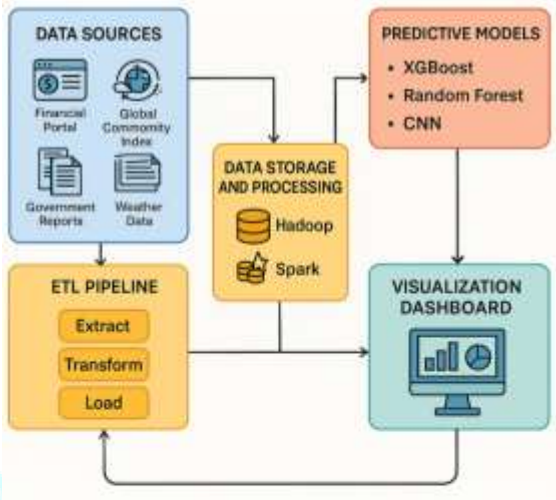


Figure 1: Proposed coffee commodity trading System

METHODOLOGIES

The methodology for this project follows a structured pipeline that integrates data engineering, machine learning, and user interface design to deliver a scalable and intelligent coffee trading system.

1. Data Collection: The system collects coffee commodity data from various reliable sources including Yahoo Finance, Commitment of Traders (COT) reports, World Bank datasets, and custom web scraping tools. These sources provide real-time and historical data on coffee prices, trade volumes, and market behavior.

2. ETL Pipeline Implementation: An automated ETL (Extract, Transform, Load) process is implemented in three layers. In the extraction phase, structured, semi-structured, and unstructured data is gathered. The transformation layer involves data cleaning, handling of missing values, normalization, and feature engineering (e.g., moving averages, volatility indices). Finally, the data is loaded into a multi-layered architecture consisting of a staging area, operational data store, and data marts optimized for analytical queries.

3. Data Storage and Processing: Apache Hadoop is used for scalable storage, while Apache Spark handles real-time distributed data processing. These technologies enable high-speed handling of large datasets and ensure fault tolerance and flexibility in operations.

4. Predictive Modeling: Several models are trained, including Random Forest, XGBoost, and Convolutional Neural Networks (CNN). The XGBoost model is fine-tuned using hyperparameter optimization to reduce error rates. A voting ensemble mechanism combines predictions from different models to ensure robustness and accuracy.

5. Visualization and Interaction: The system includes an interactive web-based interface built with Django. Visualization dashboards provide dynamic insights into coffee trading patterns, enabling users to explore trends and forecast prices interactively.

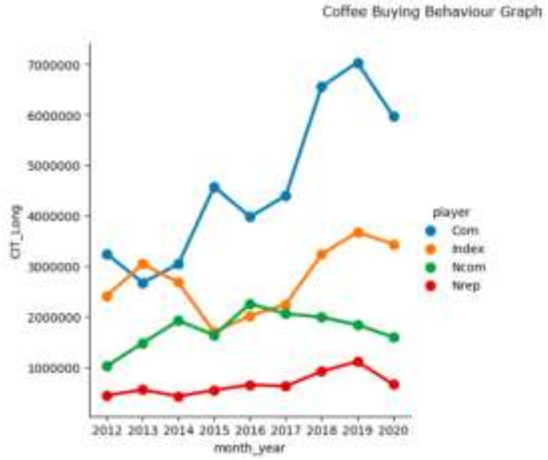
RESULTS

The developed system was evaluated using real-world coffee trading datasets sourced from platforms like Yahoo Finance and the Commodity Futures Trading Commission (CFTC). The system demonstrated strong performance in predicting coffee price trends and visualizing market behavior. Key results include a significant reduction in prediction error when using the optimized XGBoost model, with a Root Mean Square Error (RMSE) of 1.12 compared to 2.45 in traditional models.

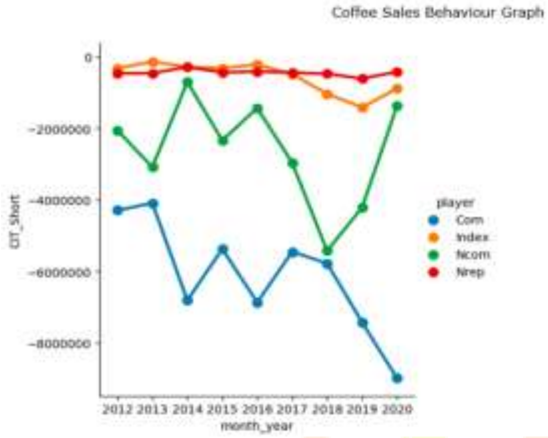
The ensemble voting mechanism further enhanced prediction accuracy by combining outputs from multiple models, including Random Forest and CNN. This approach provided more consistent forecasts, even during periods of high market volatility.

Real-time visualization dashboards enabled users to track historical trends, analyze buyer-seller behavior, and observe shifts in coffee prices across different years. These insights helped users make more informed investment decisions.

Additionally, the system's real-time data ingestion and processing capabilities ensured that updated market trends were quickly reflected in the analytics platform. The web-based user interface was tested for responsiveness, usability, and accuracy of outputs, receiving positive feedback from trial users.



Buy Analysis



Sales Analysis

Algorithm Name	MSE Score	RMSE Score	Model Performance
Propose Random Forest	0.018317204647094697	0.13534106785116887	Worst Model
Extension Tuned XGBoost	0.0013039556527845197	0.036110326124039915	Best Model
Propose Deep Learning CNN	0.0289252543068889	0.17007426115344115	Worst Model

ML/DL Algorithms Performance

DISCUSSION

The implementation of artificial intelligence and big data technologies in this project marks a transformative step in the domain of commodity trading. By integrating a layered ETL framework with advanced machine learning algorithms, the system addresses key limitations in traditional trading platforms—namely, limited scalability, low prediction accuracy, and lack of real-time analytics. The optimized use of XGBoost, combined with ensemble learning, ensures that predictions are not only accurate but also resilient in volatile market conditions.

One of the most impactful features is the system's real-time data processing and visualization capabilities. These dashboards allow investors to explore historical trends, identify market shifts, and anticipate future price fluctuations effectively. The visual representation of buyer-seller behavior and seasonal variations in coffee prices adds depth to the decision-making process, helping users mitigate risks and capture timely opportunities.

Despite its strengths, the system still relies heavily on the availability and quality of external data sources. Any inconsistency or lag in data feeds can affect prediction accuracy. Furthermore, while the current architecture is scalable, maintaining low latency under heavy loads remains a challenge for future optimization.

CONCLUSION

As we have seen, this project showcases the application of AI, Big Data, and sophisticated analytics in transforming coffee commodity trading. The system processes large amounts of trading data in real-time by applying a three-tier ETL pipeline and utilizing Hadoop and Spark. Reliably optimized ML models, including XGBoost and ensemble voting, provided remarkable accuracy in forecasting and strong reliability in decision making. By depicting essential market patterns, interactions, and future prices, interactive dashboards increase user engagement. Not only does the framework solve the inadequacies of current systems, like insufficient scalability and real-time visibility, but it also equips investors with robust, real-time tools for informed risk management and return optimization. Even with persistent issues around data accuracy and timeliness, the system as a whole offers a modern financial market solution that is intelligent, scalable, and adaptable. This framework paves the way for the use of AI technology in supporting decision-making in the rapidly changing commodity markets.

REFERENCES

[1] N. Hubballi and P. Khandait, ‘‘KeyClass: Efficient keyword matching for network traffic classification,’’ Comput. Commun., vol. 185, pp. 79–91, Mar. 2022, doi: 10.1016/j.comcom.2021.12.021.

[2] W. Wei, H. Gu, W. Deng, Z. Xiao, and X. Ren, ‘‘ABL-TC: A lightweight design for network traffic classification empowered by deep learning,’’ Neurocomputing, vol. 489, no. 7, pp. 333–344, Jun. 2022, doi: 10.1016/j.neucom.2022.03.007.

[3] H. Han, Z. Yan, X. Jing, and W. Pedrycz, ‘‘Applications of sketches in network traffic measurement: A survey,’’ Inf. Fusion, vol. 82, pp. 58–85, Jun. 2022, doi: 10.1016/j.inffus.2021.12.007.

[4] C. Do, D. Duong, and D. Hoang, ‘‘A multi-layer approach for advanced persistent threat detection using machine learning based on network traffic,’’ J. Intell. Fuzzy Syst., vol. 40, no. 6, pp. 11311–11329, Jan. 2021, doi: 10.3233/JIFS-202465.

[5] P. A. Mitroshin, Y. Y. Shitova, Y. A. Shitov, A. A. Mitroshin, and D. N. Vlasov, ‘‘GIS-monitoring of regional transport network traffic as a method to study commuting: Moscow region case,’’ J. Phys., Conf. Ser., vol. 1828, no. 1, Feb. 2021, Art. no. 012073, doi: 10.1088/1742-6596/1828/1/012073.

[6] C. D. Xuan, H. Thanh, and N. T. Lam, ‘‘Optimization of network traffic anomaly detection using machine learning,’’ Int. J. Electr. Comput. Eng., vol. 11, no. 3, pp. 2360–2370, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2360-2370.

[7] C. D. Xuan, ‘‘Detecting APT attacks based on network traffic using machine learning,’’ J. Web Eng., vol. 20, no. 1, pp. 171–190, Jan. 2021, doi: 10.13052/jwe1540-9589.2019.

[8] L. Duan, J. Zhou, Y. Wu, and W. Xu, ‘‘A novel and highly efficient botnet detection algorithm based on network traffic analysis of smart systems,’’ Int. J. Distrib. Sensor Netw., vol. 18, no. 3, pp. 182459–182476, Mar. 2022, doi: 10.1177/15501477211049910.

[9] Z. Long and W. Jinsong, ‘‘Network traffic classification based on a deep learning approach using NetFlow data,’’

Comput. J., vol. 66, no. 8, pp. 1882–1892, Aug. 2023, doi: 10.1093/comjnl/bxac049.

[10] V. Ponnusamy, A. Yichiet, N. Jhanjhi, M. Humayun, and M. F. Almufareh, ‘‘IoT wireless intrusion detection and network traffic analysis,’’ Comput. Syst. Sci. Eng., vol. 40, no. 3, pp. 865–879, Mar. 2022, doi: 10.32604/csse.2022.018801.

[11] L. Li, V. Okoth, and S. E. Jabari, ‘‘Backpressure control with estimated queue lengths for urban network traffic,’’ IET Intell. Transp. Syst., vol. 15, no. 2, pp. 320–330, Feb. 2021, doi: 10.1049/itr2.12027.

[12] T. Zanma, D. Hashimoto, K. Koiwa, and K. Liu, ‘‘Estimation of network traffic status and switching control of networked control systems with data dropout,’’ IET Cyber-Phys. Syst., Theory Appl., vol. 7, no. 2, pp. 69–80, Nov. 2021, doi: 10.1049/cps2.12024.

[13] C. Wang, H. Zhou, Z. Hao, S. Hu, J. Li, X. Zhang, B. Jiang, and X. Chen, ‘‘Network traffic analysis over clustering-based collective anomaly detection,’’ Comput. Netw., vol. 205, Mar. 2022, Art. no. 108760, doi: 10.1016/j.comnet.2022.108760.

[14] S. Dong and Y. Xia, ‘‘Network traffic identification in packet sampling environment,’’ Digit.Commun.Netw., vol. 9, no. 4, pp. 957–970, Aug. 2023, doi: 10.1016/j.dcan.2022.02.003.

[15] G. Sri Vidhya and R. Nagarajan, ‘‘Performance analysis of network traffic intrusion detection system using machine learning technique,’’ Int. J. Commun. Antenna Propag., vol. 12, no. 2, p. 111, Apr. 2022, doi: 10.15866/irecap.v12i2.21724.

[16] Y. Peng, W. Wu, J. Ren, and X. Yu, ‘‘Novel GCN model using dense connection and attention mechanism for text classification,’’ Neural Process.Lett., vol. 56, no. 2, pp. 1–17, Apr. 2024, doi: 10.1007/s11063-024-11599- 9.

[17] M. Zulqarnain, R. Ghazali, M. Aamir, and Y. M. M. Hassim, ‘‘An efficient two-state GRU based on feature attention mechanism for sentiment analysis,’’ Multimedia Tools Appl., vol. 83, no. 1, pp. 3085–3110, Jan. 2024, doi: 10.1007/s11042-022-13339-4.

[18] M. Laurer, W. van Atteveldt, A. Casas, and K. Welbers, ‘‘Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI,’’ Political Anal., vol. 32, no. 1, pp. 84–100, Jan. 2024, doi: 10.1017/pan.2023.20.

[19] F. Ullah, S. Ullah, G. Srivastava, and J. C.-W. Lin, ‘‘IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic,’’ Digit. Commun.Netw., vol. 10, no. 1, pp. 190–204, Feb. 2024, doi: 10.1016/j.dcan.2023.03.008.

[20] A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, ‘‘Advancing multilingual handwritten numeral recognition with attention-driven transfer learning,’’ IEEE Access, vol. 12, pp. 41381–41395, 2024, doi: 10.1109/ACCESS.2024.3378598.