

¹Dinesh Singh
Dhakar²Dr. Ritu Jain³Dr. Anjali Dadhich⁴Roshni Rajput

Investigating the Role of Machine Learning and Blockchain for Forecasting Stock Market Trends



Abstract: - There is a sudden deluge of evolutionary approaches in the domain of data processing and computation, which are significantly affecting several facets of applications globally. Artificial intelligence, machine learning and Blockchain happen to be at the forefront of evolutionary computation beating conventional approaches. Finance applications currently are heavily reliant of data driven models. Stock trend analysis happens to be one such approach, which lays the foundation for forecasting decisions to be made. The leeway in such applications is critically small as minimal inaccuracies in forecasting may lead to major losses. This paper presents the current perspective in terms of evolutionary algorithms such as artificial intelligence and Blockchain and how they are transforming software development, the application of machine learning algorithms to regression problems. Finally, the stock trend analysis based on in and out of sample datasets has been performed for standard S&P datasets. A comparative analysis with previous work clearly indicates the improved performance of the proposed work with respect to baseline approaches in the domain.

Keywords: Evolutionary Algorithms, Machine Learning, Blockchain, Regression Problems, Stock Trend Analysis.

I. INTRODUCTION

Financial The software industry is witnessing a complete overhaul in terms of the conceptualization, clientele, functioning, development, resource management and testing. Moreover, recent trends such as remote work ad-hoc global teams, virtual teams and virtual software development rely heavily on technologies such as AI and Blockchain (Salah et al., 2019). Covid1-19 fast-tracked the need and adoption of global teams for software development, making global software development an imminent necessity. Ad-hoc and virtual teams, remote jobs and the integration of automation for software development are now being used extensively to cater to a global market to attain cost-reduction, maintain product quality, retaining company vision and culture, enhancing collaboration among global teams while attracting the best talent pool available globally. The use of AI, Machine Learning and Deep Learning tools such as ChatGPT and MemGPT have infused large scale automation in software development, which has resulted in a two-fold scenario.

While such tools have made global software development easier, with virtual assistants, interactive development and auto generation, it has also resulted in job losses to automation (Pham et al, 2022). Thus skill displacement and management for developers worldwide needs to be considered seriously. Blockchain on the other hand, focusses on de-centralized computation especially related to digital transactions and interconnectivity. This is particularly important for Web 3.0 applications which needs to ensure transparency and traceability for both homogenous and heterogeneous networks and software. Thus, AI, ML and Blockchain are the most sought after technologies and skillsets for global software development (Liu et al., 2022).

¹ New Horizon Institute of Technology and Management, Thane, India.

Saraswati College of Engineering, Kharghar, Navi Mumbai, India.^{2,3,4}

dineshhodcsemit@gmail.com¹

ritujain.mca@gmail.com², anjalidadhich01@gmail.com³, rajput.rosh89@gmail.com⁴

Corresponding Author: Dinesh Singh Dhakar: dineshhodcsemit@gmail.com

Copyright © JES 2024 on-line : journal.esrgroups.org

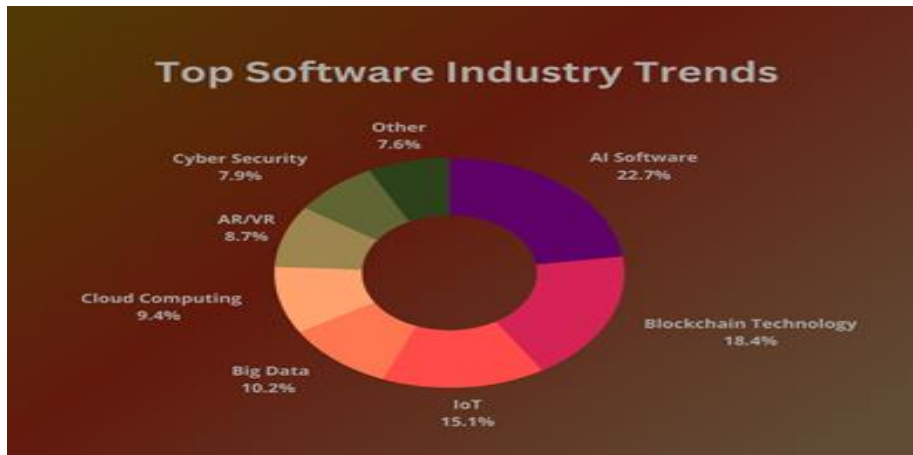


Figure 1: Top Industry Trends

(Source: www.goodfirms.com/resources/software-development-research)

Figure 1 depicts the recent software industry trends, which clearly shows that AI Software and Blockchain happen to have the maximum share of 22.7% and 18.4% respectively making them the most important emerging technology stacks. IoT and Big Data happen to be the close 3rd and 4th which also happen harness AI and Blockchain technologies. Moreover, global software development is witnessing a paradigm shift with the emergence and widespread application of emerging technologies such as artificial intelligence (AI) and Blockchain. While AI and Blockchain are making their presence felt in several domains, it is making a significant impact on the software development lifecycle (SDLC), starting from conceptualization, concurrent development, managing resources, deployment and testing. A metamorphosis is also being witnessed with large scale automation and distributed computing, making conventional approaches non-existent while creating newer opportunities. This paper focusses on the emergence of emerging technologies and its impact on software development as a whole, with a focus on global software development trends being impacted by the same. Multiple use cases pertaining to use of AI and Blockchain in global software development have been cited and analyzed. It is expected that the paper would render significant insight into the working, salient features, applications limitations and concerns of such new age technologies and their influence of global software development and management.

II. DATA DRIVEN MODELS AFFECTING SOFTWARE DEVELOPMENT

Both AI and Blockchain are playing a significant role in it. global software development are being revolutionized by AI/ML and Blockchain technologies. Some of the recent trends are (Lato et al., 2023):

1. Use of Automation in Software Development: The entire software development lifecycle (SDLC) is being transformed by AI/ML and Blockchain by enhancing productivity through virtual assistants and auto generation of code (Ernst et al., 2022).
2. AI driven technologies such as Agile and DevOps are aiding collaboration, customer feedback, and the ability to respond changes with low latency (Shaw et al., 2022).
3. AI driven ready to ship code is being auto generated which saves a lot of time for software developers and also the production time.
4. Reliability and testing are also being automated reducing latencies and enhancing production quality and reliability.

Auto generation and debugging have become much easier with AI tools along with code maintenance and quality assurance. Moreover AI tools are being used to integrate APIs (Shaw et al., 2022). AI integrated APIs can handle much more requests across multiple platforms compared to conventional APIs. From a developers perspective, the actual functionality of the AI/ML algorithms invoked in the backend are not mandatory to be known in detail. Additionally, AI driven virtual assistants can be utilized to act as middleware between humans and API. AI enabled prediction tools can also enable developers to estimate the performance of software. Furthermore, AI enabled tools can come in handy in analyzing large code bases and documentation, interpreting complex systems and developing new software (Ramchand et al., 2021).

Blockchain technology is characterized by decentralized, secure, transparent and traceable computation. These attributes are key in enhancing the performance of the SDLC along with enhancing the security of distributed software across multiple platforms (Bankar et al., 2021). Software supply chain and pipeline are also employing Blockchain to enhance robustness and security. Additionally data integrity which is a crucial factor for global software development can be bolstered through Blockchain as recorded inputs can not be modified (Wei et al., 2021). This combining AI/ML and Blockchain can enhance the productivity of global software development as well as making them more secure through distributed and decentralized networks (Benito et al., 2021).

III. DATA DRIVEN MOELS FOR FINANCIAL AND STOCK FORECASTING

AI is revolutionizing stock investments by introducing innovative approaches to analysis, decision-making, and portfolio management. Here's a breakdown of how AI is transforming the stock investment landscape: AI excels in handling vast amounts of financial data, analyzing historical trends, and identifying complex patterns. Machine learning algorithms can recognize subtle market signals and patterns that may be challenging for human analysts to discern. This enables investors to make more informed decisions based on comprehensive data analysis. AI models use predictive analytics to forecast stock prices and market trends. These models analyze historical data, market indicators, and various external factors to make predictions about future price movements. This assists investors in making proactive investment decisions and managing risk more effectively (Li et al., 2021).

AI-driven algorithmic trading systems execute trades at high speeds based on predefined criteria. These algorithms can respond to market changes in real-time, optimizing trade execution and minimizing human intervention. This approach helps investors capitalize on fleeting opportunities and ensures timely reactions to market fluctuations. Natural Language Processing (NLP) is employed to analyze news articles, social media, and financial reports to gauge market sentiment. By understanding public sentiment, AI can predict market reactions and investor behavior. Investors can use this information to adjust their strategies and make decisions that align with prevailing market sentiment (Wang et al., 2020). AI plays a crucial role in assessing and managing investment risks. Machine learning models can evaluate portfolio diversification, identify potential risks, and suggest adjustments to minimize downside exposure. This proactive risk management contributes to the overall stability of investment portfolios (Thavaneswaran et al., 2020)

AI algorithms optimize investment portfolios by considering various factors such as risk tolerance, investment goals, and market conditions. These systems can suggest well-balanced portfolios that align with an investor's preferences, ultimately maximizing returns while managing risk. AI-powered robo-advisors provide personalized investment advice based on individual financial goals, risk tolerance, and market conditions. This democratizes access to financial guidance, making it more accessible to a broader range of investors. AI is employed to detect and prevent fraudulent activities in the financial markets. Advanced algorithms can analyze trading patterns and detect anomalies that may indicate market manipulation or other illicit activities, enhancing overall market integrity. AI systems continuously learn from market data and adapt to changing conditions. This adaptability ensures that investment strategies remain relevant and effective in dynamic market environments.

3.1 Machine Learning for Stock Forecasting: A Case Study

Forecasting stock market trends using machine learning models has become increasingly popular due to their ability to analyze large datasets and identify complex patterns. The major steps included are: As machine learning algorithms can analyse vast datasets and find intricate patterns, they are becoming more and more popular for forecasting stock market movements. The main components and models utilised in applying machine learning to stock market trend forecasting are broken down as follows (Shivhare et al., 2022):

Preparing data: Gathering and preparing historical market data is the initial stage in developing a stock market forecasting model. Stock prices, trading volumes, economic indicators, and pertinent news are a few examples of this data. In order to guarantee accuracy and consistency in ensuing studies, cleaning and normalising the data is essential. Stock price data is treated as a time series in several stock market forecasting models. Data trends, seasonality, and temporal patterns can all be found in time series analysis. Methods like seasonal breakdown of time series (STL) and autoregressive integrated moving average (ARIMA) are frequently employed in this setting. For stock market forecasting, a variety of machine learning techniques are utilised. Among the well-known ones are:

Stock prices are predicted using linear regression, which is based on linear relationships with particular features. Decision Trees: Models based on trees are capable of capturing intricate relationships seen in data. Random Forests: Decision tree ensembles offer increased robustness and accuracy. Support Vector Machines (SVM): Assigns various market trends to individual data points.

Time series forecasting can benefit from the use of recurrent neural networks (RNNs), which are specialised for sequential data. Forecasting accuracy can be increased overall by combining forecasts from numerous models using ensemble techniques like bagging or stacking. By utilising the advantages of various algorithms, ensemble methods make up for the shortcomings of individual algorithms (Sisodiya et al., 2022).

3.2 Regression Learning

The regression learning approach needs to fit the data over a wide set of ranges which can be:

- 1) In sample.
- 2) Out of sample.

Moreover, the case on imbalanced datasets need to be explored for the purpose as well. With long, mid and short term forecasting problems, it is essential to design an algorithm which can render high accuracy forecasting for all spans. Validating the machine learning model's performance on unobserved data is essential once it has been trained. Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are examples of common evaluation metrics. Techniques for cross-validation aid in ensuring the generalizability of the model. The efficiency of stock market forecasting models is anticipated to increase as machine learning and data availability continue to progress, giving investors important insights for making wise decisions in the intricate and ever-changing financial landscape.

IV. PROPOSED METHODOLOGY

The proposed methodology tries to implement 2 major approaches:

- 1) Pre-Process data for removing noise and disturbance effects.
- 2) Trend analysis for a wide range of stock data.

The Median Filter

The median filter is normally used to reduce fluctuations in data. It has a property of preserving useful details in the data. The median filtering is applied by taking the median value for a set of samples with an experimentally determined width as follows:

$$y_{med}(n) = \mathbf{median}(x[n:n + k]) \quad (1)$$

Here,

$y_{med}(n)$ is the median filter

(n) is the number of data samples

k is the window size to apply the filter.

The median filter has low computational complexity as opposed to other stochastic filters and hence suitable for large data samples such as crypto-prices.

wavelet transform on the independent variables and then trains the neural network with the values.

The BFGS/Quasi Newton Algorithm

The Quasi Newton Back Propagation often termed as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

The essence of the algorithm is the fact that it is fact and works better than most 2^{nd} order gradient descent

algorithms. While most 2nd order gradient descent algorithms have a complexity of $O(N^2)$, the BFGS algorithm has a complexity of $O(N)$ thereby drastically reducing the computational complexity. It computes the Hessian Matrix \mathbf{H} , as a more effective approximation. The Hessian Matrix is defined as:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 e}{\partial x_1 \partial w_1} & \dots & \frac{\partial^2 e}{\partial x_1 \partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 e}{\partial x_n \partial w_1} & \dots & \frac{\partial^2 e}{\partial x_n \partial w_n} \end{bmatrix} \quad (2)$$

The 2nd order derivative based learning algorithms essentially compute the inverse of the Hessian Matrix to update the weights of the network in each iterations, given by:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha [\mathbf{H}]^{-1} \frac{\partial e}{\partial \mathbf{w}} \quad (3)$$

Here,

k and $k + 1$ denotes the present and next iterations of training.

w_k & w_{k+1} denote the weights of the present and subsequent iterations.

α denotes the learning rate.

e denotes the error in the present iterations.

$\frac{\partial e}{\partial w}$ is called the error gradient

The proposed algorithm essentially presents the deep neural network model with data optimization, which is expressed as a sequential implementation of the following steps:

Step.1 Extract dataset and divide data into the ratio of 70:30 for training : testing.

Step.2 Apply data filtering employing median filter.

Step.3 Initialize weights randomly.

Step.4 Initialize training through the BFGS algorithms training rule:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha [\mathbf{H}]^{-1} \frac{\partial e}{\partial \mathbf{w}}$$

Step.5 If (cost function stabilizes)

Truncate training

Else if (max. iterations are over)

Truncate Training

Else

Feedback errors as inputs to subsequent iteration.

Step.7 if (error is stable through validation checks i.e. 6 consecutive iterations)

Stop training

else if (maximum iterations are over even without error stabilization)

Stop Training

else

{

```

Feed next training vector
Back propagation of error
}

```

Step.8 Compute performance metrics.

V. EXPERIMENTAL RESULTS

The experimental results have been obtained for benchmark S& P datasets. We start out analysis with the Amazon dataset. The same approach translates to all other datasets.



Fig.2 Raw Amazon Dataset



Figure 3 Marked Statistical Features of the data.

Table 1: Statistical feature of raw data

S.No.	Parameter	Value
1	Samples	2517
2	Min	14.35
3	Max	186.6
4	Mean	80.37
5	Median	82.75
6	Mode	57.71

The statistical features of the data are presented in table 1. The forecasting results for long, mid and short term forecasts are presented next.

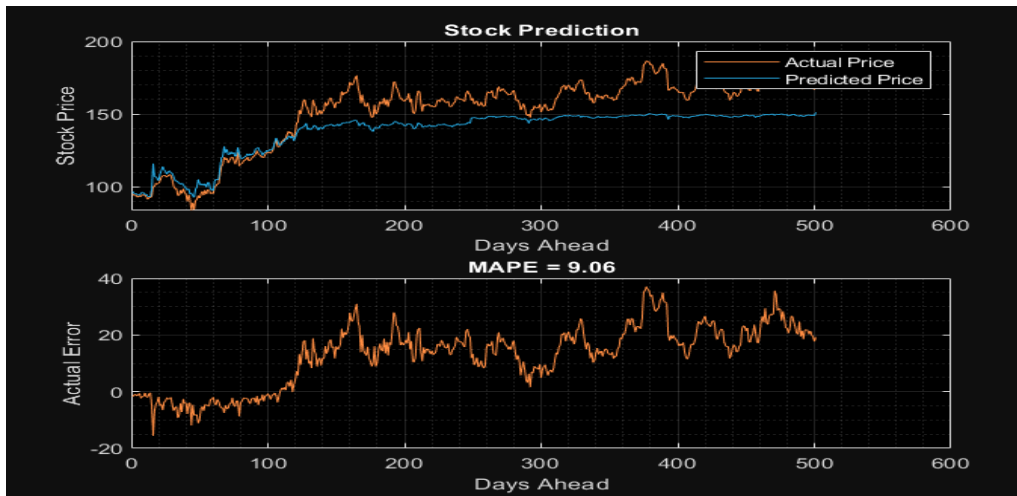


Figure 4 Long Term Forecast

Figure 4 depicts the long term forecast results over a period of 500 days. It can be observed that the proposed approach attains an MAPE of 9.06%. The next part would be forecasting the mid and short term forecasts and computing the error rates and accuracy.

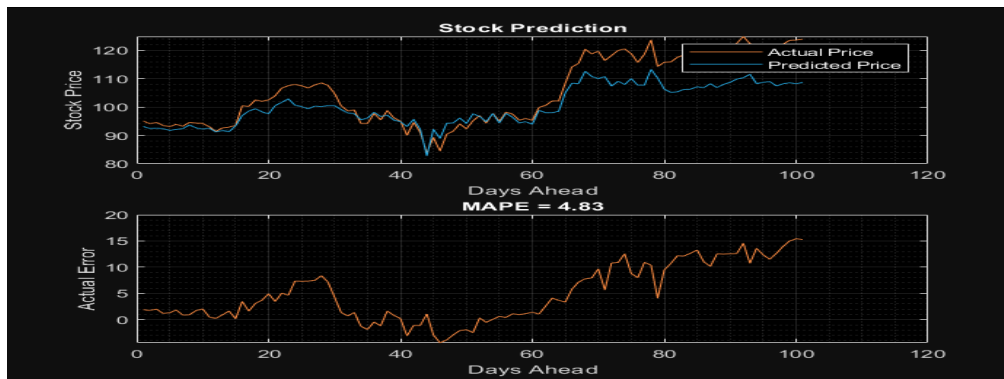


Figure 5 Mid term forecast

The mid term forecast has been done for a period of 100 days (over 3 months). The accuracy achieved is 4.83%

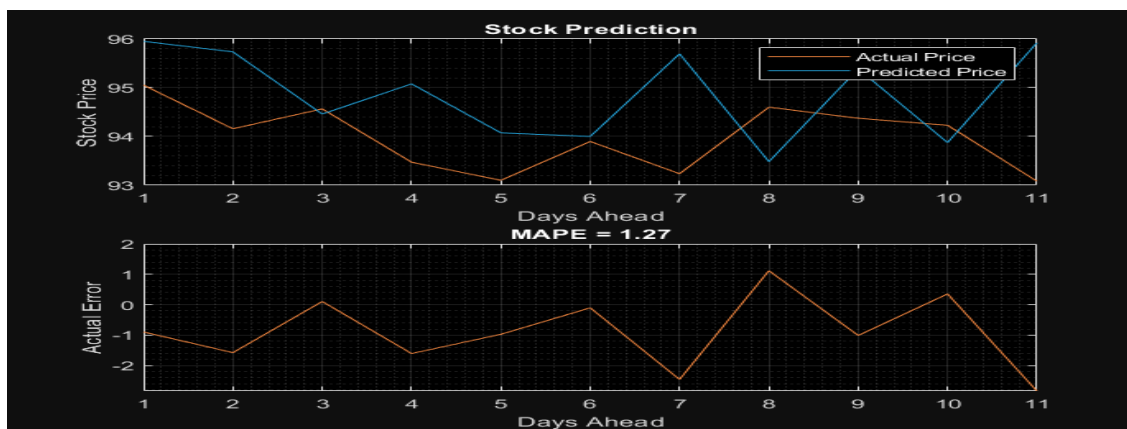


Figure 6 Short term forecast

The short term forecast has been done for a period of 10 days (over 3 months). The accuracy achieved is 1.27% only

Table 2. MAPE Comparison

S.No.	Duration	Days Ahead	MAPE
1	Long Term	500	9.06%
2	Mid Term	100	4.83%
3	Short Term	10	1.27%
4.	Average MAPE%		5.053%
5.	LSTM-ARO		6.58%
	LSTM-GA		7.583%
	LSTM1D		8.889%
	LSTM2D		8.659%
	LSTM3D		8.784%

It can be observed that the proposed approach attains a sub 10% MAPE for all 3 forms of forecast i.e. long, mid and short. However, the highest MAPE% happens to be in the long term forecast as the amount of variability to be encountered needs to be very high over a period of almost 2-10 years. The mid term forecast MAPE of 4.83% is somewhat in between that of the long and short term forecasts. Shorter duration forecasts are typically more accurate as the stock market needs time to change trends, barring exceptional cases (crashes or surges). The average MAPE % for the approach across all terms is 5.053%. A comparative analysis with the most recent trends based on LSTM models (Gülmez et al.) is also compared in table 2. Five models based on the LSTM model have been compared with this work. The five models are the Artificial Rabbit Optimization based LSTM (ARO-LSTM), genetic algorithm based LSTM (GA-LSTM), LSTM1D, LSTM2D and LSTM3D models which have 1, 2 and 3 LSTM layers respectively. It can be observed that the proposed approach outperforms all the models in terms of forecasting accuracy.

VI. CONCLUSION

This paper presents a modern overview of evolutionary algorithms for both trend analysis and software development. A case study of stock trend analysis based on deep neural networks has been presented. It can be concluded that stock trend analysis is complex in nature due to the variability in the dataset parameters and hence needs the trend analysis prowess of machine learning and data driven models. The proposed model presents a pre-processing approach along with deep neural networks for stock trend analysis. The analysis tenures have been chosen as long, short and mid to cover all investing modalities (temporal approach). It can be observed that the proposed approach attains a forecasting accuracy of 5.053% (mean) outperforming baseline approaches in the domain.

Author contributions

Dinesh Singh Dhakar, Dr. Ritu Jain: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation., Field study. **Dr. Anjali Dadhich, Roshni Rajput:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

- [1] K. Salah, M. H. U. Rehman, N. Nizamuddin and A. Al-Fuqaha, "Blockchain for AI: Review and Open Research Challenges," in *IEEE Access* 2019, vol. 7, pp. 10127-10149.
- [2] P Pham, V Nguyen, T Nguyen, A Review of AI-augmented End-to-End Test Automation Tools, ASE '22: Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, 2022, Art.214, pp.1-4.
- [3] D Liu, Y Zhang, D Jia, Q Zhang, X Zhao, "Toward secure distributed data storage with error locating in blockchain enabled edge computing", *Computer Standards & Interfaces*, Elsevier 2022, vol.79., 103560.

- [4] S Laato, M Mäntymäki, AKMN Islam, Trends and Trajectories in the Software Industry: implications for the future of work, *Information System Frontiers*, Springer 2023, vol.25, pp.929-944.
- [5] N. A. Ernst and G. Bavota, "AI-Driven Development Is Here: Should You Worry?," in *IEEE Software* 2022, vol. 39, no. 2, pp. 106-110.
- [6] R. Kapur and B. Sodhi, "Towards a Knowledge Warehouse and Expert System for the Automation of SDLC Tasks," 2019 IEEE/ACM International Conference on Software and System Processes (ICSSP), Montreal, QC, Canada, 2019, pp. 5-8.
- [7] M. Shaw and L. Zhu, "Can Software Engineering Harness the Benefits of Advanced AI?," in *IEEE Software*, 2022, vol. 39, no. 6, pp. 99-104.
- [8] S Ramchand, S Shaikh, I Alam, "Role of artificial intelligence in software quality assurance", *Proceedings in Intelligent Systems and Applications. IntelliSys*, Springer 2021, vol.295.
- [9] Conference on Global Software Engineering, 2014, pp. 153-162
- [10] KV Siakas, E Siakas, "The need for trust relationships to enable successful virtual team collaboration in software outsourcing", *International Journal of Technology, Policy and Management*, Inderscience 2008, vol.8, no.1, pp. 59-75.
- [11] J Cruz-Benito, S Vishwakarma, F Martin-Fernandez, "Automated source code generation and auto-completion using deep learning: Comparing and discussing current language model-related approaches", *AI*, MDPI, 2021, vol.2, no.1, pp.1-16.
- [12] G Giray, "A software engineering perspective on engineering machine learning systems: State of the art and challenges", *Journal of Systems and Software*, Elsevier, 2021, vol.180, 111031.
- [13] J Zarrin, H Wen Phang, L Babu Saheer, B Zarrin, "Blockchain for decentralization of internet: prospects, trends, and challenges", *Cluster Computing*, Springer 2021, vol.24., pp. 2841–2866.
- [14] S Demi, R Colomo-Palacios, M Sanchez-Gordon, "Software engineering applications enabled by blockchain technology: A systematic mapping study", *Applied Sciences*, MDPI, 2021, vol.11, no.7, 2960.
- [15] S. Bankar and D. Shah, "Blockchain based framework for Software Development using DevOps," 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), NaviMumbai, India, 2021, pp. 1-6
- [16] PC Wei, D Wang, Y Zhao, SKS Tyagi, "Blockchain data-based cloud data integrity protection mechanism", *Future Generation Computer Systems*, Elsevier 2020, vol.102, pp. 902-911.
- [17] X. -H. Li et al., "A Survey of Data-Driven and Knowledge-Aware eXplainable AI," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 29-49, 1 Jan. 2022.
- [18] Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," in *China Communications*, vol. 17, no. 3, pp. 205-221, March 2020.
- [19] A. Thavaneswaran, Y. Liang, Z. Zhu and R. K. Thulasiram, "Novel Data-Driven Fuzzy Algorithmic Volatility Forecasting Models with Applications to Algorithmic Trading," 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, 2020, pp. 1-8.
- [20] B. D. Shivahare, S. Suman, S. S. N. Challapalli, P. Kaushik, A. D. Gupta and V. Bibhu, "Survey Paper: Comparative Study of Machine Learning Techniques and its Recent Applications," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 449-454.
- [21] P. S. Sisodia, A. Gupta, Y. Kumar and G. K. Ameta, "Stock Market Analysis and Prediction for Nifty50 using LSTM Deep Learning Approach," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 156-161.
- [22] B Gülmez, "Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm", *Expert System with Applications*, Elsevier, 2023, vol.29. 120346.