Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

The Emergence of Evolutionary Algorithms and its Applications to Regression Problems: A Stock Trend Analysis Case Study

[1]Dr. Harmaninder Jit Singh, [2]Er. Amanpreet Singh, [3]Durdana Taranum Khan, [4]Dr. Nithinsha Shajahan

[1]Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India

[2]Asst Professor

University college Benra-dhuri (sangrur)

Department of Computer Science and Engineering, Desh Bhagat University, India

[3]College of Computer Science & Information Systems

Department of Information Technology & Security, Jazan University

[4]Lecturer Department of Information Technology and Security.

College of Computer Science and Information Technology

Jazan University, KSA

***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.

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

1. Introduction

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).

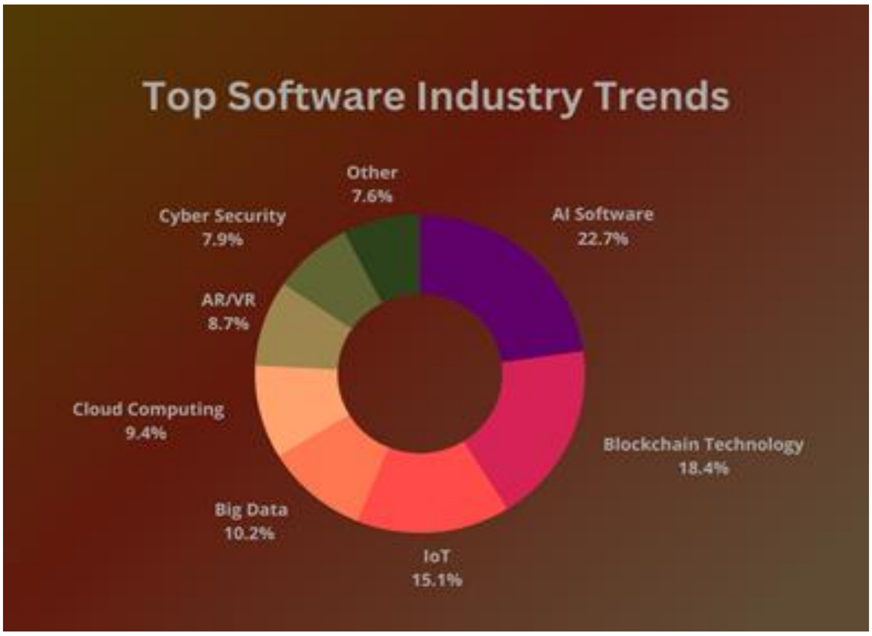

Fig.1: Top Industry Trends

(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.

2. 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 developer's 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). Blockchain is also employed by pipeline and Software supply chain in order to augment security as well as strength. In addition to that, there is a crucial factor called data integrity, in case of comprehensive software development that can be boosted with the help of Blockchain because the logged inputs are immutable (Wei et al., 2021). The combination of Blockchain as well as AI/ML can lead to the improvement in the production of worldwide development of software. It also ensures that with the help of decentralized and distributed networks they are protected to greater extent (Benito et al., 2021).

3. Data Driven Moels For Financial And Stock Forecasting

Stock investments are being revolutionized by artificial intelligence. This is being made possible by presenting state-of-the-art methods to portfolio management, scrutiny, policymaking etc. This is the split-up of the scenario in which the artificial intelligence is transmuting the scene of stock investment: It is helpful in tackling massive amount of financial data. It is equally good in scrutinizing recognizing multifaceted patterns and past trends. ML algorithms can identify delicate market signs and forms. For a human analyst these parameters may be difficult to determine. On the basis of this wide-ranging data scrutiny investors are able to make more well-versed decisions. Models based on artificial intelligence make use of prognostic analytics to predict market trends as well as stock prices. This happens because these models examine market indicators, various exterior factors, historic data etc. in order to predict future price trends. This enables the investors to make more practical decisions about investment. They are able to manage risk more effectually (Li et al., 2021).

AI driven algorithms can reply to market changes very effectively and in real-time. So, they are able to optimize trade accomplishment and also minimize the human involvement. This slant helps investors to ensure timely responses to market variations. Natural Language Processing is used to investigate news articles. It is also helpful in analyzing financial reports and social media. Last but not the least it also measures market mawkishness. By, AI can foresee market responses and investor's conduct by making use of public sentimentality. Investors use this info to regulate their approaches. They go for conclusions in accordance with the ongoing market sentiment (Wang et al., 2020). A vital role is played by AI in measuring and handling investment risks. Machine learning (ML) models can handle portfolio divergence, recognize perils, and can recommend measures to diminish downside revelation. This risk management is helpful in constancy of investment portfolios (Thavaneswaran.et al., 2020)

3.1 Machine Learning for Stock Forecasting: A Case Study

Making use of ML models in foretelling stock market inclinations is becoming more and more popular. This is due to the fact that they have capability to examine bulky datasets. They can also identify intricate patterns.

Following is the breakup the foremost models and components that are involved in application of ML to stock market trend prediction (Shivhare et al., 2022):

Preparing data: The initial step in collecting and making historical market data is the development of the stock market prediction model. A number of examples of this kind of data are - transaction volumes, relevant news, stock prices, economic gauges etc. To ensure correctness and reliability in data, cleaning and normalization of data is utmost important. A number of stock market prediction models treat data of stock prices as time series. In time series analysis the time-based patterns, seasonality, data tendencies etc. may be found. There are methods like autoregressive integrated moving average (ARIMA) and seasonal breakdown of time series that are commonly employed. For stock market forecasting, there are diversified techniques which are available in case of machine learning that are being utilized. One such well-known technique is - linear regression. Stock prices are foretold using this technique. This method/technique is based on linear relationships among specific features.

Decision Trees: Models which are based on trees have the capability of apprehending complicated relationships noticed in the data.

Random Forests: Decision trees provide enhanced precision and robustness.

Support Vector Machines (SVM): It allocates different market trends to distinct data points.

Time series prediction can be benefitted by the use of recurrent neural networks (RNNs). They are particular for sequential data. The overall forecasting accuracy can be augmented by compounding different models' forecasts. This can happen by making use of collaborative techniques like stacking or bagging. Individual algorithms' inadequacies can be overcome by making use of advantages of different algorithms collectively (Sisodiya et al., 2022).

3.2 Regression Learning

The regression learning method is required to fit the data across a wide range that can be:

1) In sample.

2) Out of sample.

The issue of imbalanced datasets is also needed to be explored. In case of short, mid and long term forecasting glitches, it is indispensable to develop an algorithm that requires more accurate forecasting for all extents. When the data has been trained, it is required to validate the ML model's performance on the basis of unobserved data. We need to have common evaluation metrics in this case e.g. Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). 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.

4. 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:

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

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 **2nd 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 **H,** as a more effective approximation. The Hessian Matrix is defined as:

$$H = \left[\frac{\partial^2 e}{\partial x1 \partial w1} \cdots \frac{\partial^2 e}{\partial x1 \partial wn} \ \vdots \ddots \vdots \vdots \ \frac{\partial^2 e}{\partial xn \partial w1} \cdots \vdots \ \frac{\partial^2 e}{\partial x1 \partial wn}\right] \quad \textbf{(2)}$$

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

$$w_{k+1} = w_k - \alpha [H]^{-1} \frac{\partial e}{\partial w} \quad \textbf{(3)}$$

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.

αdenotes the learning rate.

$e$denotes the error in the present iterations.

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

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:*

$$w_{k+1} = w_k - \alpha [H]^{-1} \frac{\partial e}{\partial w}$$

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.*

5. 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

Fig.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.

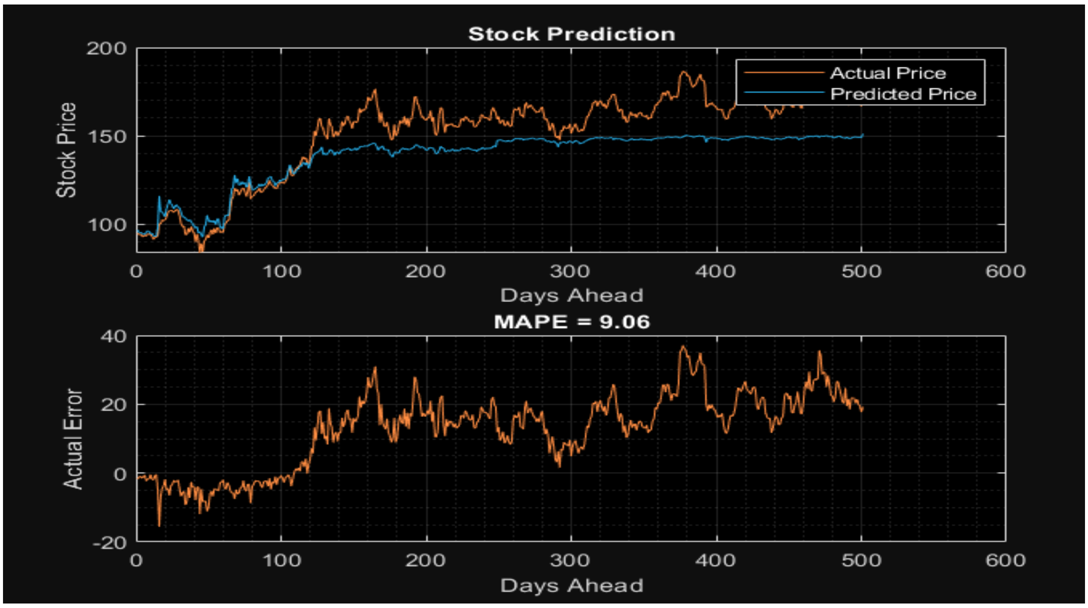

Fig.4Long Term Forecast

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.

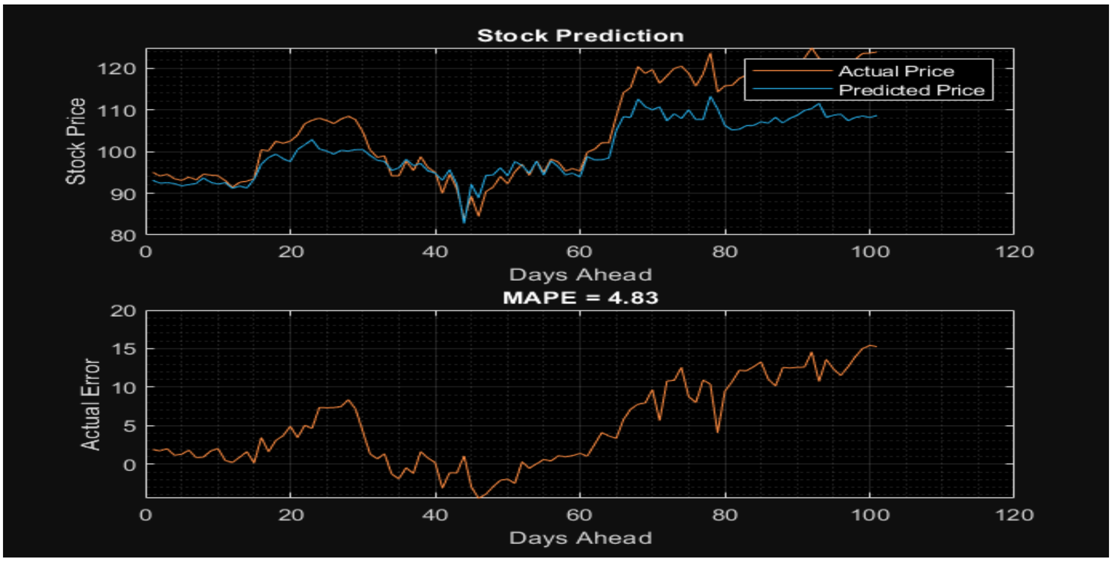

Fig.5Mid term forecast

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

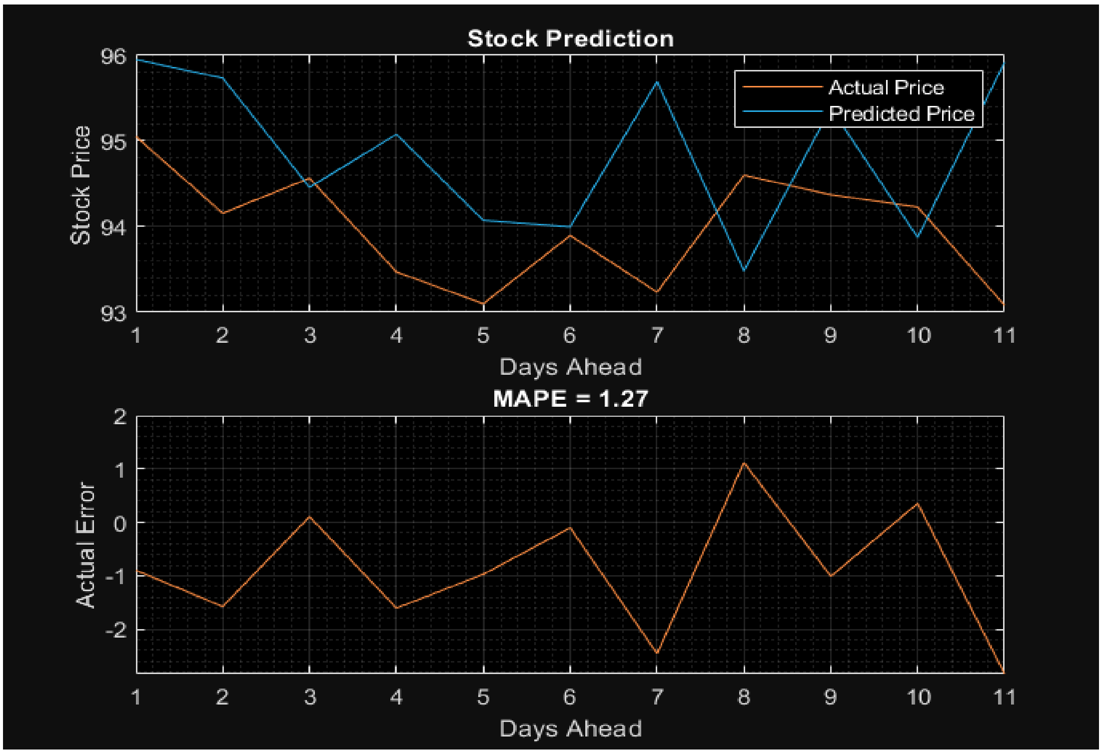

Fig.6 Short term forecast

The short-term forecast has bene 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%

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).

Conclusion

This paper presents a modern overview of evolutionary algorithms for both trend analysis and software development. A case study of stock rend 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 9.06%, 4.83% and 1.27% for the benchmark Amazon Dataset.

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] Choudhary, G., Mendiratta, H. and Kaur, D., 2023, July. Invasive Ductal Carcinoma Prediction in Mammography and Histopathology Images Using CNN. In 2023 World Conference on Communication & Computing (WCONF) (pp. 1-6). IEEE.

[23] Kukreja, V., Sharma, R. and Vats, S., 2023, August. A Hybrid Deep Learning Approach for Multi-Classification of Heritage Monuments Using a Real-Phase Image Dataset. In 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 29-32). IEEE.