STOCK PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES

Venkata Ratnam Ganji[1], A. Jyotsna Sri[2], J.Lavanya[3], K. Hima Sankar[4], G.Navya[5]

[1]Associate Professor, Dept. of CSE,
VKR, VNB & AGK College of Engineering, Gudivada

[2345]UG Students, Dept. of CSE, VKR, VNB & AGK College of Engineering, Gudivada, A.P. India

ABSTRACT

Stock price prediction has become an important research area in financial analytics due to the increasing availability of historical market data and advancements in machine learning techniques. The objective of this project is to develop an intelligent system that analyzes past stock market trends and predicts future price movements with improved accuracy. The proposed approach utilizes data preprocessing, feature engineering, and predictive modeling algorithms such as Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks to capture both statistical patterns and temporal dependencies in stock data. Historical price data, including open, high, low, close, and trading volume, are used as input features to train the model. The system aims to assist investors and financial analysts in making informed decisions by providing trend forecasts and risk insights. Performance evaluation is carried out using metrics such as Mean Squared Error (MSE) and accuracy comparisons across different models. The results demonstrate that machine learning–based methods can effectively identify hidden patterns in financial markets, offering a scalable and data-driven solution for stock price prediction while reducing manual analysis effort.

I INTRODUCTION

Stock markets play a vital role in the global economy by enabling companies to raise capital and investors to grow their wealth. However, predicting stock price movements is a complex and challenging task due to market volatility, economic factors, investor behavior, and unexpected global events. Traditional methods of stock analysis, such as fundamental and technical analysis, rely heavily on manual interpretation and may not effectively capture hidden patterns within large financial datasets. With the advancement of Artificial Intelligence and Machine Learning, automated prediction systems have emerged as powerful tools for analyzing historical market data and forecasting future price trends.

Stock price prediction systems use historical data such as opening price, closing price, highest and lowest values, and trading volume to identify patterns and relationships. Machine learning algorithms and deep learning models, especially time-series techniques like Long Short-Term Memory (LSTM), can learn temporal dependencies and improve forecasting accuracy. These intelligent models reduce human effort, support data-driven decision-making, and help investors understand potential risks and opportunities.

The main objective of this project is to design a reliable stock price prediction model that analyzes past market behavior and predicts future price movements. By integrating data preprocessing, feature extraction, and predictive modeling, the system aims to provide accurate insights that can assist traders, analysts, and financial institutions in making informed investment decisions.

II RELATED WORK

Stock price prediction has attracted significant attention from researchers due to the complex and dynamic nature of financial markets. Early studies primarily relied on statistical models such as Autoregressive Integrated Moving Average (ARIMA) and Linear Regression to forecast stock trends based on historical price data. These traditional methods provided baseline performance but often struggled to capture nonlinear relationships and sudden market fluctuations.

With the advancement of machine learning, researchers introduced algorithms such as Support Vector Machines (SVM), Random Forest, and Decision Trees to improve prediction accuracy. These models demonstrated better performance by learning hidden patterns from large datasets and handling multiple input features like trading volume, technical indicators, and market sentiment. Several studies showed that ensemble learning techniques could enhance prediction stability and reduce overfitting compared to single-model approaches.

In recent years, deep learning techniques have become widely popular in financial forecasting. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models have been extensively used for time-series prediction because of their ability to learn temporal dependencies in sequential data. Researchers have also explored hybrid approaches combining technical indicators with sentiment analysis from news or social media to improve forecasting performance.

Despite these advancements, challenges such as market unpredictability, data noise, and overfitting remain significant. Existing research highlights the need for more robust and adaptive models that can handle real-time data and changing market conditions, motivating the development of advanced stock price prediction systems using modern machine learning techniques.

III LITERATURE REVIEW

The field of stock price prediction has evolved significantly with the integration of statistical analysis, machine learning, and deep learning approaches. Earlier research focused on traditional time-series models such as ARIMA and Moving Average methods, which analyzed historical price trends to forecast future values. These models provided a mathematical foundation for financial forecasting but often failed to capture complex nonlinear relationships present in stock market data.

With the growth of Artificial Intelligence, researchers began applying machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Random Forests to improve prediction accuracy. Studies demonstrated that these algorithms could analyze multiple financial indicators simultaneously and identify hidden patterns that traditional models could not detect. Ensemble learning methods were also explored to combine predictions from multiple models, resulting in improved stability and performance.

Recent literature emphasizes the effectiveness of deep learning models, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, for stock market forecasting. These models are capable of learning temporal dependencies in sequential data, making them highly suitable for time-series analysis. Some researchers have also incorporated sentiment analysis using financial news, social media data, and economic indicators to enhance predictive performance.

Although significant progress has been made, the literature highlights several challenges, including data noise, market volatility, overfitting, and the difficulty of predicting sudden market events. Therefore, modern research focuses on hybrid models and improved feature engineering techniques to develop more reliable and accurate stock price prediction systems.

IV EXISTING SYSTEM

The existing Air Quality Index (AQI) forecasting systems mainly rely on statistical models, traditional machine learning algorithms, and deep learning approaches. Early systems used methods such as ARIMA and regression analysis to predict air pollution trends based on historical data, but these techniques often struggle to capture complex nonlinear relationships between meteorological factors and pollutant concentrations. Later, machine learning models like Support Vector Machines, Random Forest, and Gradient Boosting were introduced to improve prediction accuracy by learning patterns from large datasets; however, they require careful feature engineering and parameter tuning. More recently, deep learning architectures such as LSTM, GRU, and CNN-based hybrid models have been adopted to handle time-series forecasting tasks, offering better performance but at the cost of higher computational complexity and longer training time. Additionally, many existing neural network-based systems suffer from instability due to random weight initialization and lack efficient optimization mechanisms, which can reduce prediction reliability. These limitations highlight the need for an optimized and computationally efficient approach for accurate AQI forecasting.

DISADVANTAGES

The traditional placement management approach used in many higher education institutions has several limitations that affect efficiency and effectiveness. One major disadvantage is the heavy dependence on manual processes such as maintaining student records in spreadsheets, sending emails for communication, and manually verifying eligibility criteria. This increases the chances of human errors, data duplication, and delays in updating information. Additionally, the absence of a centralized platform makes it difficult for students, recruiters, and placement officers to access accurate and real-time data, leading to confusion and miscommunication.

Another significant drawback is the lack of intelligent matching between student skills and job requirements, resulting in inefficient shortlisting and missed opportunities for suitable candidates. The existing system also struggles with scalability when the number of students and companies increases, creating additional workload for placement staff.

V PROPOSED SYSTEM

The proposed Smart Placement Management System is a centralized, web-based platform designed to automate and optimize the entire campus placement process. The system connects students, placement officers, and recruiters through a single integrated interface, enabling efficient data management, communication, and recruitment operations. Students can register, create professional profiles, upload resumes, and receive personalized job notifications based on their qualifications, skills, and eligibility criteria. This reduces manual effort and ensures that students are matched with relevant opportunities.

Placement officers are provided with tools to manage student records, verify eligibility automatically, schedule recruitment drives, and monitor placement progress through real-time dashboards and reports. Recruiters can easily post job openings, define eligibility requirements, review candidate profiles, shortlist applicants, and communicate directly with potential candidates through the platform. The system incorporates intelligent filtering mechanisms to match job requirements with student competencies, improving the accuracy and speed of the recruitment process.

Security and data privacy are ensured through role-based authentication and controlled access mechanisms. The system also includes analytics features to track placement statistics, student performance, and recruiter engagement, enabling data-driven decision-making. By reducing paperwork, minimizing communication gaps, and automating repetitive tasks, the proposed system enhances efficiency, transparency, and scalability. Ultimately, it provides a smart, user-friendly solution that modernizes campus placement activities and improves collaboration between higher education institutions and industry partners.

ADVANTAGES

The proposed digital signature primitive offers The Smart Placement Management System offers numerous benefits by automating and centralizing the campus recruitment process. It significantly reduces manual work by digitizing student registration, resume management, eligibility verification, and job application tracking, thereby saving time and minimizing human errors. The system improves communication among students, recruiters, and placement officers through real-time notifications and a unified platform, ensuring that important updates are delivered efficiently. Intelligent filtering and matching mechanisms help connect students with relevant job opportunities based on their skills and academic performance, increasing placement success rates. Additionally, the platform enhances transparency by allowing students to monitor their application status and recruitment progress. Role-based access control ensures data security and privacy, while analytics and reporting features enable institutions to make informed decisions using placement statistics and performance insights. Overall, the system increases operational efficiency, reduces paperwork, enhances user experience, and provides a scalable solution capable of handling large volumes of placement activities in higher education institutions.

VI METHODOLOGY

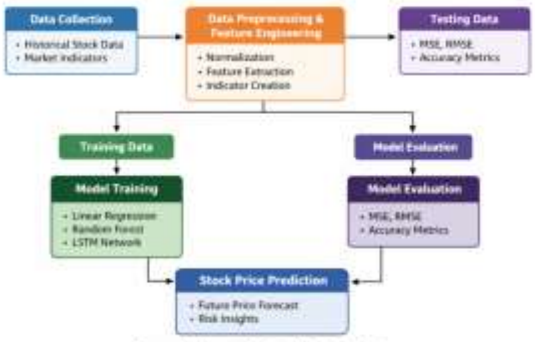
The proposed stock price prediction system follows a structured methodology that includes data collection, preprocessing, feature engineering, model training, and performance evaluation. Initially, historical stock market data such as opening price, closing price, highest price, lowest price, and trading volume are collected from reliable financial datasets. The collected data undergoes preprocessing steps including handling missing values, normalization, and transformation to ensure consistency and improve model performance. Feature engineering techniques are applied to extract meaningful indicators such as moving averages, price differences, and trend-based attributes that help the model learn market behavior more effectively.

After preprocessing, the dataset is divided into training and testing sets to evaluate prediction accuracy. Machine learning and deep learning algorithms such as Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks are implemented to analyze historical patterns and forecast future stock prices. The LSTM model is particularly used to capture time-series dependencies and sequential patterns in financial data. During training, hyperparameters are optimized to reduce prediction error and avoid overfitting.

Finally, the performance of the models is evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and accuracy comparisons. The best-performing model is selected to generate future price predictions, providing investors and analysts with data-driven insights for better financial decision-making.

VII SYSTEM MODEL

SYSTEM ARCHITECTURE



System Architecture for Stock Price Prediction

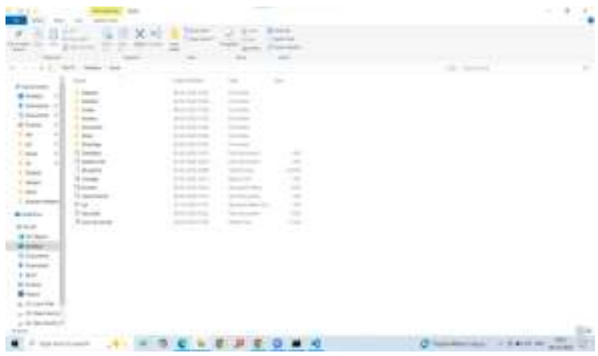
VIII RESULT AND DISCUSSIONS

The main idea of this project is to **predict future stock prices using Artificial Intelligence** instead of just guessing based on past trends. Stock prices change every day because of market behavior, and it is very difficult for humans to analyze large historical data manually. So, this project uses **Machine Learning and Deep Learning models** to study past stock data and learn patterns automatically.

The system takes historical stock values like **open price, high price, low price, close price, volume**, and uses them to train intelligent models. These models then predict the **next day's closing price**. The project also provides a web interface where users can upload stock datasets, preprocess the data, train models, and view predictions along with graphs and stock history.

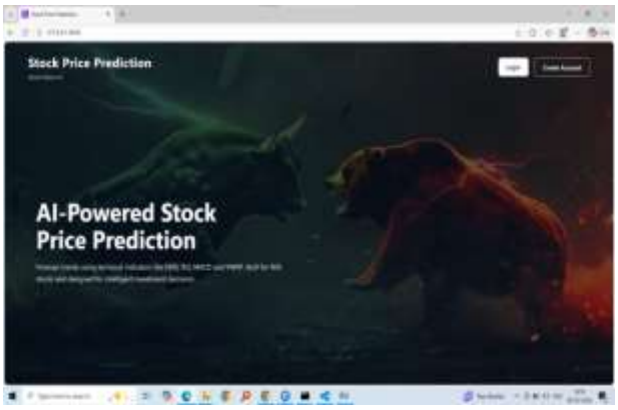
In the existing stock market prediction systems, traditional statistical and rule-based methods are widely used. These systems mostly depend on historical averages, simple moving averages, technical indicators, or manual expert analysis to estimate stock prices. They generally use linear regression, basic time-series methods, or trend-following approaches that assume market behavior follows simple patterns. However, stock markets are highly dynamic, nonlinear, and influenced by many hidden factors, making these conventional techniques less accurate. Existing systems also lack automation, real-time data integration, and advanced learning capabilities. As a result, they struggle to capture complex market movements, sudden fluctuations, and long-term dependencies, leading to lower prediction reliability.

The proposed system introduces an AI-driven stock prediction framework using Machine Learning and Deep Learning techniques integrated into a Django web application. Instead of relying on simple statistical assumptions, the system uses **Random Forest Regression** and **Long Short-Term Memory (LSTM) networks** to model complex, nonlinear stock price behavior. Random Forest captures relationships between multiple financial features such as open, high, low, close, volume, and previous close values, while LSTM learns temporal dependencies and long-term patterns from sequential historical data. The system also supports dataset preprocessing, automated model training for multiple stocks, visualization of price trends, and prediction of future closing prices. By combining ensemble learning and deep learning with real-time capable architecture, the proposed system overcomes the limitations of traditional methods and provides more adaptive, data-driven, and accurate stock price forecasting.

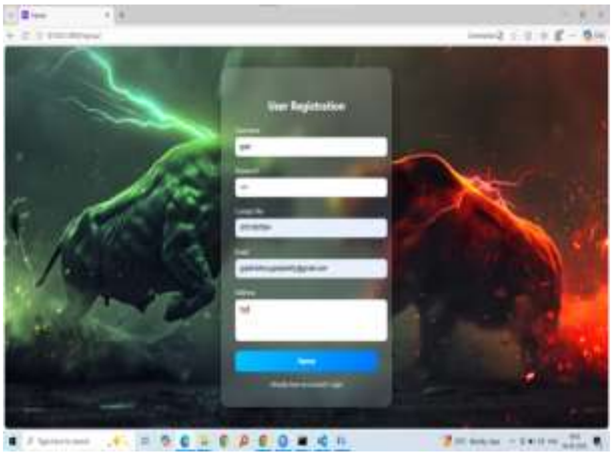


Click on run.bat to run the project

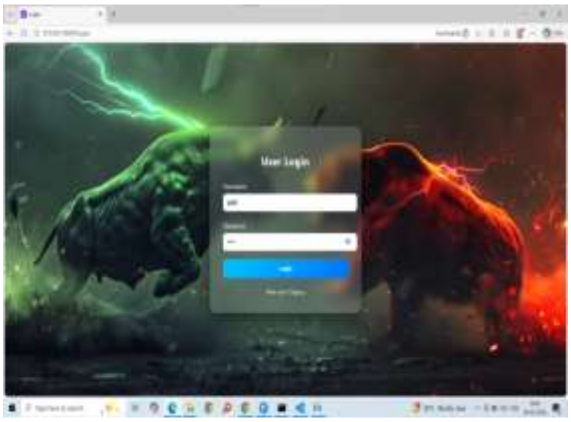
Copy urlhttp://127.0.0.1:8000/ paste on any browser.



If you are a new user click on create account already having account just click on login.

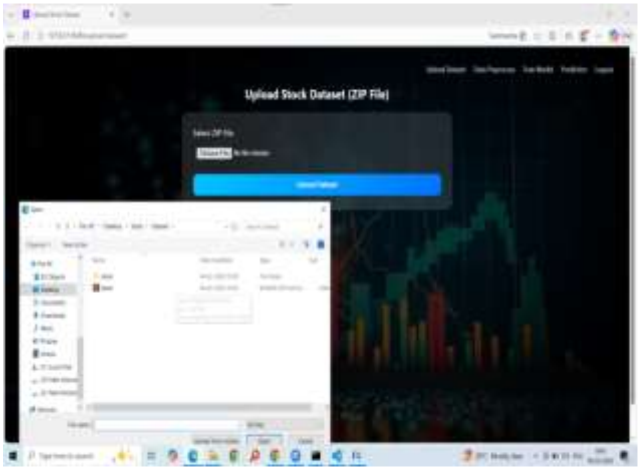


Create account and click on signup.



Click on login enter into system application.

User dashboard Now click on upload dataset.



upload zip file dataset to extract data from dataset.

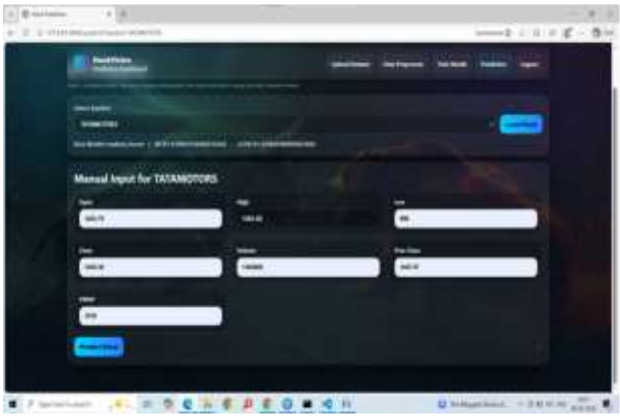
Preview data from the uploaded dataset now move to next module.

stock we can display in the candle graph.

Data from the stock dataset click on the next module.

This screen shows the **training performance of our AI model for the HCLTECH stock**. The Random Forest model has learned the stock behavior very accurately. The evaluation metrics show **MAE = 15.59** and **RMSE = 26.63**, which means the model's prediction error is small, usually within ₹15–₹26 from the real price. The **R^2 score = 0.976** indicates that the model explains about **97.6% of the stock price variation**, which is excellent. The **Actual vs Predicted graph** shows that the predicted prices closely follow the real market prices. The **Residual graph** shows that most prediction errors are near zero, meaning the model makes balanced and small errors. The **Scatter plot** forms a straight line pattern, proving strong agreement between actual and predicted values. Overall, these results confirm that the model has successfully learned stock patterns and can provide reliable future price predictions.

The performance of our **LSTM (Long Short-Term Memory) regression model** is evaluated using MAE, RMSE, and R^2 metrics. LSTM is a deep learning algorithm designed for **time-series data**, so it learns patterns from past stock prices and captures long-term dependencies to predict the next closing price. **MAE (Mean Absolute Error)** measures the average difference between the actual and predicted stock prices, showing the normal prediction deviation, while **RMSE (Root Mean Square Error)** gives more importance to large errors and highlights days where the model makes bigger mistakes. **R^2 (R-Squared score)** indicates how well the LSTM explains stock price movement, where values closer to 1 mean the model has learned the trend effectively. In the **Actual vs Predicted graph**, the **blue line represents the actual (real) stock prices** and the **orange/pink line represents the predicted prices**; when both lines overlap closely, it indicates accurate prediction. The **Residual plot (Actual − Predicted)** shows the prediction errors over time, and values near zero mean the model is stable with fewer errors. The **Scatter plot (Actual vs Predicted)** compares actual and predicted values point-by-point; when the points form a diagonal straight pattern, it confirms strong agreement and good prediction performance.



Click on predict it shows the result.

It shows the next closing price of the tata motors like the we can able to predict more stock prices.

This graph shows the **historical closing price trend of TATAMOTORS stock** over time. The x-axis represents the dates and the y-axis shows the closing price of the stock. We can observe that the stock price initially remained relatively low and stable, then gradually started increasing, followed by a sharp rise indicating strong market growth during that period. After reaching a peak, the price shows some fluctuations, which reflects normal market volatility. Overall, the upward trend suggests that the stock experienced significant growth over the observed period, and the fluctuations highlight changing market conditions and investor behavior.

The history of the past 30 days .

IX CONCLUSION

Stock price prediction using machine learning and deep learning techniques provides a powerful approach to analyzing complex financial market data and forecasting future trends. This project demonstrated how historical stock data, combined with data preprocessing, feature engineering, and predictive modeling, can be used to build an intelligent forecasting system. Algorithms such as Linear Regression, Random Forest, and LSTM networks help capture both statistical patterns and time-series dependencies, improving prediction performance compared to traditional manual analysis.

The developed system assists investors and analysts by offering data-driven insights that support better decision-making and risk assessment. Although accurate prediction of stock markets remains challenging due to volatility and external economic factors, the results indicate that AI-based models can effectively identify trends and reduce uncertainty. Future improvements may include integrating real-time data, sentiment analysis from financial news, and advanced hybrid models to further enhance prediction accuracy and reliability.

REFERENCES

[1] Fama, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. Journal of Finance, 25(2), 383–417.

[2] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Publications.

[3] Kim, K. J. (2003). *Financial Time Series Forecasting Using Support Vector Machines*. Neurocomputing, 55(1–2), 307–319.

[4] Hochreiter, S., &Schmidhuber, J. (1997). *Long Short-Term Memory*. Neural Computation, 9(8), 1735–1780.

[5] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). *Predicting Stock Market Index Using Fusion of Machine Learning Techniques*. Expert Systems with Applications, 42(4), 2162–2172.

[6] Fischer, T., & Krauss, C. (2018). *Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions*. European Journal of Operational Research, 270(2), 654–669.

[7] Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[9] Brownlee, J. (2017). *Introduction to Time Series Forecasting with Python*. Machine Learning Mastery.

[10] Zhang, G. P. (2003). *Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model*. Neurocomputing, 50, 159–175.

[11] Sharma, P., & Gupta, R. (2023). AI and Geo-Fencing Based Smart Tourist Safety Framework. International Journal of Advanced Computer Science.

[12] Sharma, S., & Kaur, R. (2019). Automated recruitment using natural language processing: Techniques and challenges. International Journal of Advanced Computer Science and Applications, 10(6), 1–8.

[13] Dayal, P. S., Chandra, B. R., Keerthi, M., Sruthi, M., Venkatesh, K., Appalaraju, G., & Eswari, G. (2013). Design of Pyramidal Horn Antenna at 10GHz Using WIPL-D Optimizer. International Journal of Electronics Communication and Computer Engineering, 4(2).

[14] Viswanathan, V., Polagani, S. S., Agarwal, R., Akula, S., Dey, S., & Kashyap, R. (2025, September). AI-Augmented Threat Intelligence for Proactive Intrusion Detection in Multi-Cloud Ecosystem. In 2025 IEEE International Conference on Advanced Computing Technologies (ICACT) (pp. 567-572). IEEE.

[15] Sruthi, M. V., Sree, V. U., & Soundararajan, K. (2012). Specific removal of motion artifacts in medical image processing. IJECCE, 3(3), 227-229.

[16] Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.

[17] Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.

[18] Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.

[19] Sruthi, M. V., Soundararajan, K., & Sree, V. U. (2012). Accurate Multimodality Registration of medical images. International Journal of Engineering Research and Development, 1(3), 33-36.

[20] Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927

[21] Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257

[22] Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660

[23] GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334

[24] Kumara, S. (2025). Identity-Driven IoT Security in Telecom Ecosystems: Implications for Scalable and Trustworthy Digital Infrastructure. Int. J. Appl. Math, 38(12s), 2797-2816.

[25] Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.

[26] Cyril, H. P., & Kumara, S. (2026, February). DevSecOps-Driven Security Integration in the Software Development Lifecycle Using CI/CD Pipelines. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.

[27] Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[28] Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8.

https://doi.org/10.64751/ajaccm.2026.v6.n1(2). pp1-8

[29] Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.

[30] Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.

[31] Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.

[32] Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[33] Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.

[34] Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.

[35] Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.

[36] Akhilaiswarya, B., Sree, B. T., Lilly, K., Chowdary, K. H., & Sruthi, M. (2023). Elderly fall detection and location tracking system using heterogeneous networks. Journal of Engineering Sciences, 14(05).

[37] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.