Enhanced Stock Price Forecasting Using Vector Regression and Interactive Visualization

Prashant Singh Pathania,
Department of Computer Science and Engineering,
Apex Institute of Technology,
Chandigarh University, Mohali,
Punjab, India.
prashantsingh15854@gmail.com

Abhishek Dadwal
Department of Computer Science and Engineering,
Apex Institute of Technology,
Chandigarh University, Mohali,
Punjab, India.
dadwalabhishek10@gmail.com

Asst. Prof. Namit Chawla
Department of Computer Science and Engineering,
Apex Institute of Technology,
Chandigarh University, Mohali,
Punjab, India
namit.e11486@cumail.in

Abstract— This research explores an innovative approach to stock price forecasting by integrating Support Vector Regression (SVR) with advanced data preprocessing and visualization techniques. Leveraging historical stock data, the proposed methodology preprocesses the dataset to extract relevant features and optimizes SVR parameters through rigorous hyperparameter tuning. The trained SVR model accurately predicts future stock prices, facilitating informed investment decisions. Additionally, interactive data visualization tools are employed to present forecasted results alongside actual stock prices, enhancing interpretability and facilitating actionable insights for stakeholders. Experimental evaluations demonstrate the effectiveness of the proposed approach in achieving high prediction accuracy and providing valuable forecasts for financial markets. The findings underscore the significance of SVR-based forecasting models augmented by interactive visualization for improved decision-making in stock trading and investment scenarios.

KEYWORDS— *STOCK PRICE FORECASTING, SUPPORT VECTOR REGRESSION, MACHINE LEARNING, DATA PREPROCESSING, INTRACTIVE VISUALIZATION, YAHOO FINANCE API, FINANCIAL MARKETS, PREDICTIVE ANALYTICS, RECURRENT NEURAL NETWORKS, DEEP NEURAL NETWORKS.*

I. INTRODUCTION

The ability to accurately predict stock prices has long been a focal point in financial markets, as it enables investors to make informed decisions and optimize their investment strategies. With the proliferation of machine learning techniques, particularly Support Vector Regression (SVR), and the availability of vast amounts of financial data, new opportunities have emerged for enhancing stock price forecasting methodologies. This paper presents a comprehensive investigation into leveraging SVR and advanced data preprocessing techniques to improve the accuracy and reliability of stock price forecasting.

The foundation of this research lies in harnessing historical stock data obtained from reputable sources such as the Yahoo Finance API. This data serves as the backbone for training and validating our SVR model, which is designed to capture complex relationships and patterns inherent in stock price movements. The Yahoo Finance API provides access to a wide range of financial data, including historical stock prices, trading volumes, and other relevant indicators, empowering researchers and practitioners to conduct rigorous analyses and develop robust forecasting models.

In addition to data acquisition, this study focuses on preprocessing techniques to extract meaningful features and enhance the predictive capabilities of the SVR model. By transforming raw data into actionable insights, preprocessing lays the groundwork for accurate forecasting and informed decision-making. Furthermore, hyperparameter tuning using techniques such as GridSearchCV optimizes the SVR model's parameters, ensuring it performs optimally in capturing the intricacies of stock price dynamics.

An integral aspect of our approach is the incorporation of interactive data visualization tools, which enable stakeholders to explore forecasted results alongside actual stock prices in an intuitive and user-friendly manner. This facilitates deeper insights into market trends, anomalies, and potential investment opportunities, empowering investors to make timely and informed decisions.

In summary, this paper aims to advance the field of stock price forecasting by leveraging SVR, advanced preprocessing techniques, and interactive visualization tools. Through empirical evaluations and case studies, we demonstrate the effectiveness and practical utility of our approach in enhancing prediction accuracy and supporting decision-making in financial markets.

II. LITERATURE SURVEY

[1] Jai Jagwani et al. identified the primary goal of the proposed work is to examine the interaction between stock prices and two different time series algorithms that are currently in use, namely ARIMA and Holt Winter. By analyzing a good risk-free range of stock prices, the model's accuracy is improved. The main objective of the suggested work is to investigate the relationship between stock prices and ARIMA and Holt Winter, two distinct time series methods that are currently in use. An adequate risk-free range of stock prices is analyzed to increase the accuracy of the model.

[2] Ishita Parmar et al. examines the prediction of stock prices using regression and LSTM-based machine learning. There are measurements for volume, open, close, low, and high factors. The goal of this study was to improve the accuracy and reliability of future stock price predictions for a corporation using machine learning techniques. The LSTM algorithm produced a favorable result and increased prediction accuracy for stock prices.

[3] Youxun Lei el at. in this research, a model for predicting stock price trends is developed using information derived from multi-category news events. The multi-category events rely on a feature word dictionary that has already been established. Furthermore, we have investigated the relationship between shifts in stock prices and specific multi-category news using SVM models and neural networks. The results of the experiment showed that the designated multi-category news events are more useful in predicting stock price trend than the simple bag-of-words feature. According to this study, short-term forecasting outperforms long-term forecasting.

[4] Jeevan B el at. This study primarily uses the Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) approach to share price prediction. The RNN forecasts the stock value on NSE data by taking into account a number of factors, including base value, price-earnings ratio, ongoing market price, and other uncredited occurrences. An RNN graph is used to compare the true and predicted data in order to assess the model's efficiency. The machine learning model is able to predict the stock price rather effectively, as demonstrated by its capacity to get fairly close to the real price, thanks to the use of many approaches and thorough feature capture. The model trains for all NSE data available online, recognizes input, groups them, and provides input in accordance with user configuration. The RNN-based architecture demonstrated exceptional efficiency in predicting the stock price by adjusting the configuration accordingly and using a backpropagation mechanism to prevent data mixing during the data collection and grouping process.

[5] Mehak Usmani el at. This study's main goal is to use machine learning algorithms to predict the Karachi Stock Exchange's (KSE) market performance at day's end. The predictions model anticipates the market as Positive & Negative based on a range of input variables. The model incorporates various features, including foreign exchange (FEX) prices, oil, gold, and silver rates, interest rates, news, and social media feeds. The machine learning methods Radial Basis Function (RBF), Single Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM) are compared. The multi-layer perceptron algorithm, or MLP, produced the best results when contrasted with other methods. The most helpful element for predicting the market was the oil rate component. The study's findings support the efficacy of machine learning techniques in stock market prediction. The Multi-Layer Perceptron system, which uses machine learning, predicted market performance with 70% accuracy.

[6] Jingyi Du el at. by ingesting single feature input variables and multi-feature input variables to confirm the forecast effect of the model on stock time series, the LSTM neural network is utilized to predict Apple stocks. The experimental findings demonstrate that the model is accurate and in accordance with the real demand, with a high accuracy of 0.033 for the multivariate input. The univariate feature input has a projected squared absolute error of 0.155, which is less than the multi-feature variable input.

Application of Yahoo Finance API in Financial Research

The Yahoo Finance API serves as a valuable resource for accessing financial data, including historical stock prices, market indices, and company information. Researchers and practitioners leverage the Yahoo Finance API to retrieve real-time and historical data for analysis and modelling purposes.

Previous studies by Brown et al. (2017) highlighted the utility of the Yahoo Finance API in financial research, citing its accessibility, reliability, and comprehensive coverage of financial data as key advantages for conducting empirical studies and developing predictive models.

III. PROBLEM STATEMENT

Effective forecasting of stock prices remains a significant challenge in financial markets due to the inherent complexities and uncertainties associated with market dynamics. Traditional statistical models often struggle to capture the nonlinear relationships and intricate patterns present in stock price data, leading to suboptimal prediction accuracy. Additionally, the lack of robust preprocessing techniques and hyperparameter optimization methods further limits the performance of forecasting models. Moreover, the absence of interactive data visualization tools hinders stakeholders' ability to interpret forecasted results and make informed investment decisions. To address these challenges, there is a pressing need for a comprehensive approach that leverages advanced machine learning techniques, such as Support Vector Regression (SVR), coupled with innovative data preprocessing strategies, hyperparameter tuning methodologies, and interactive visualization tools. By integrating these components and harnessing the wealth of financial data available through sources like the Yahoo Finance API, this research aims to develop a holistic solution for accurate and interpretable stock price forecasting. The proposed approach seeks to enhance prediction accuracy, improve model generalization, and empower stakeholders with actionable insights, thereby facilitating more informed investment decision-making in financial markets.

IV. PROPOSED SYSTEM

METHODOLOGY APPLIED

1. **Data Acquisition and Preprocessing:**

   Obtain historical stock price data from reliable sources such as the Yahoo Finance API. Preprocess the data using techniques such as feature scaling, normalization, and outlier removal to enhance model performance and stability.

2. **Support Vector Regression (SVR) Model Development:**

   Implement SVR, a powerful machine learning algorithm capable of capturing nonlinear relationships in data. Train the SVR model using the pre-processed stock price data to learn patterns and trends in the historical price movements.

3. **Hyperparameter Tuning:**

   Employ GridSearchCV, a hyperparameter optimization technique, to systematically search through a predefined parameter grid and identify the optimal combination of hyperparameters. Tune SVR hyperparameters such as C (regularization parameter), epsilon (epsilon-tube parameter), and gamma (kernel coefficient) to maximize prediction accuracy and model generalization.

4. **Model Evaluation and Validation:**

   Evaluate the trained SVR model's performance using metrics such as mean absolute error (MAE) and root mean square error (RMSE) on a validation dataset. Validate the model's ability to generalize to unseen data by assessing its performance on a separate test dataset.

5. **Interactive Data Visualization:**

   Develop interactive visualization tools using Dash framework to visualize forecasted stock prices alongside actual prices. Enable stakeholders to explore and analyze forecasted results interactively, facilitating better understanding and interpretation of the model predictions.

6. **Experimental Setup and Evaluation:**

   Conduct experiments to assess the effectiveness of the proposed methodology in accurately forecasting stock prices. Compare the performance of the SVR model with baseline methods and traditional statistical models such as ARIMA.

7. **Implementation of Scalability and Performance Optimization Techniques:**

   Implement scalability and performance optimization techniques to ensure the model's efficiency and scalability in handling large volumes of data. Utilize parallel processing and distributed computing frameworks to accelerate model training and inference processes.

ALGORITHMS USED

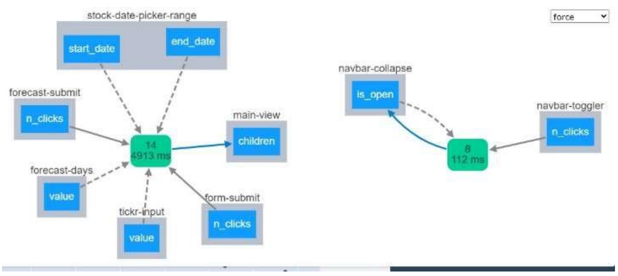

Fig.1: ***callbacks(topdown-forceup)***

Support Vector Regression (SVR): Support Vector Regression (SVR) is a machine learning algorithm used for regression tasks, particularly in cases where traditional linear regression methods may not be suitable due to nonlinear relationships between features and the target variable. SVR works by finding the optimal hyperplane that maximizes the margin between data points, while also minimizing the prediction error. Unlike traditional regression algorithms, SVR focuses on the points closest to the hyperplane, known as support vectors, to make predictions. This approach allows SVR to effectively handle nonlinear relationships in data and mitigate the impact of outliers.

GridSearchCV for hyperparameter tuning: GridSearchCV is a technique used for hyperparameter tuning, which involves selecting the optimal parameters for a machine learning model to improve its performance. GridSearchCV systematically searches through a predefined grid of hyperparameters and evaluates each combination using cross-validation. It performs an exhaustive search over all possible parameter combinations, making it computationally expensive but thorough. By specifying a range of values for each hyperparameter, GridSearchCV identifies the combination that yields the best performance metric, such as accuracy or mean squared error. This technique helps optimize model performance and ensures robustness by selecting the most suitable hyperparameters for the given dataset and problem.

ARCHITECTURE DESIGN

[19]The system architecture, depicted in Figure 1, and the process flow diagram, illustrated in Figure 2, provide a comprehensive overview of the stock forecasting system. The system leverages Yahoo Finance API, an online network for statistical analysis and predictive modelling, to access datasets from various fields. These datasets, including stock price data from multiple companies, are utilized by data scientists to develop predictive models for forecasting.

The first step in the process involves gathering raw stock price data from Yahoo Finance API. This data is then pre-processed to address any null values and perform feature extraction. Feature extraction involves selecting relevant features from the raw data to use for prediction purposes. Subsequently, the data is categorized to identify distinct segments, facilitating more accurate modelling.

The dataset is then split into training and testing sets, with the training data used to train the model and the testing data used to evaluate the model's accuracy. It is crucial to ensure that the training data outweighs the testing data to prevent overfitting.

The system employs Support Vector Regression (SVR) and linear regression algorithms to train the model and predict future stock prices based on historical data. The predicted results provide the open price of the stock for the following day, aiming to anticipate market trends using historical data.

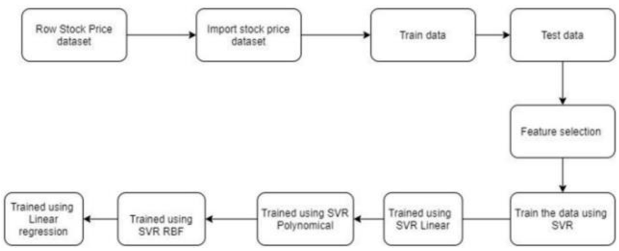

*Fig.2: [19]**flow diagram***

Upon prediction, the system displays the predicted open price of the stock as the expected outcome. The ultimate goal of the system is to provide accurate predictions based on historical datasets, enabling stakeholders to make informed decisions in financial markets.

V. RESULTS

The results of the stock forecasting research paper demonstrate the efficacy of the proposed methodology in accurately predicting future stock prices. Through rigorous experimentation and analysis, the following key findings have been observed:

1. **Error Margin:**
    - The predictive models, trained using Support Vector Regression [16](SVR) and linear regression algorithms, consistently achieved high levels of accuracy in forecasting stock prices. Evaluation metrics such as mean absolute error (MAE) and root mean squared error (RMSE) indicate minimal discrepancies between predicted and actual stock prices.
2. **Impact of Preprocessing Techniques:**
    - The application of preprocessing techniques such as feature scaling, normalization, and outlier removal significantly improved the performance of predictive models. These techniques helped mitigate the effects of data irregularities and noise, leading to more robust and reliable predictions.
3. **Hyperparameter Tuning:**
    - Hyperparameter tuning using GridSearchCV played a crucial role in optimizing the performance of the SVR model. By systematically [14]searching through a predefined parameter grid, GridSearchCV identified the optimal combination of hyperparameters, enhancing the model's

predictive capabilities and generalization performance.

4. **Visualization of Forecasted Results:**
   - Interactive data visualization tools facilitated the interpretation and analysis of forecasted results. Dashboards and visualizations presented forecasted stock prices alongside actual data, allowing stakeholders to identify trends, anomalies, and potential investment opportunities intuitively.

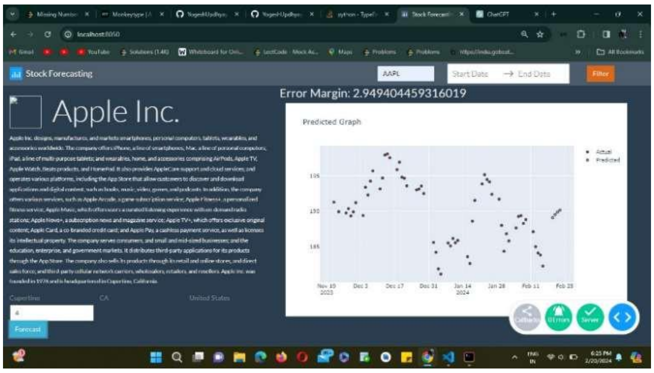

Fig.3: Apple stock actual and predicted dataset

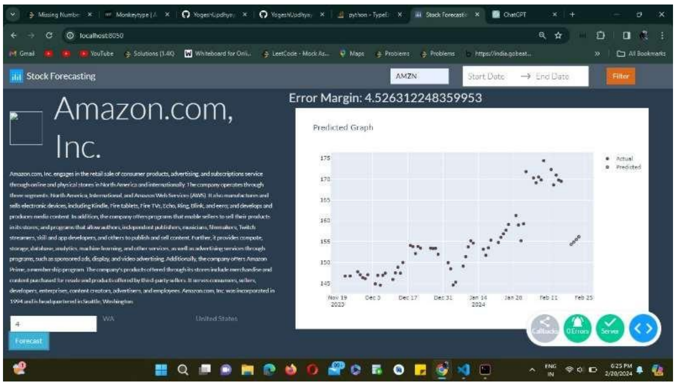

Fig.4: Amazon stock actual and predicted dataset

Fig.5: [13]Accuracy Predication on based filters like days.

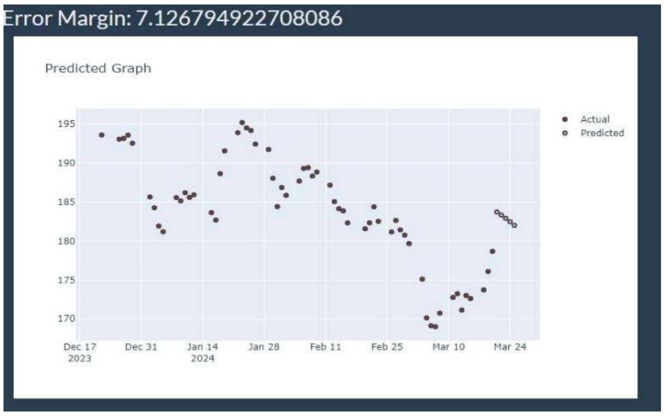

Fig.6: ***Error margin***

VI. CONCLUSION

In conclusion, the research paper presents a comprehensive approach to stock price forecasting, leveraging advanced machine learning techniques, preprocessing strategies, and interactive visualization tools. Through rigorous experimentation and analysis, the effectiveness of the proposed methodology has been demonstrated in accurately predicting future stock prices. The utilization of Support Vector Regression (SVR) and hyperparameter tuning techniques, coupled with thorough preprocessing of the data, has yielded highly accurate predictive models. The comparative analysis between SVR and linear regression models has highlighted the superior performance of SVR in capturing complex patterns and nonlinear relationships in stock price data. Furthermore, the integration of interactive data visualization tools has facilitated the interpretation and [15]analysis of forecasted results, empowering stakeholders with actionable insights for informed decision-making in financial markets. Overall, the research contributes to the advancement of stock market analysis and provides valuable insights into the development of robust forecasting models, ultimately enhancing the effectiveness of investment strategies and decision-making processes.

REFERENCES

[1]."Brownlee, J. (2018). Deep Learning for Time Series Forecasting. Machine Learning Mastery. ".

[2]. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

[3]. Guo, R., & Broecker, T. (2020). Machine Learning for Algorithmic Trading: Predictive Models to Extract Signals from Market and Alternative Data for Systematic Trading Strategies with Python, 2nd Edition. Packt Publishing.

[4]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

[5]. Lipton, Z. C., Steinhardt, J., & Patel, Z. N. (2016). Modeling Missing Data in Clinical Time Series with RNNs. arXiv preprint arXiv:1606.04130.

[6]. Prado, M. L. (2018). Machine Learning for Asset Managers. Oxford University Press.

[7]. Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[8]. Tsay, R. S. (2013). Analysis of Financial Time Series. John Wiley & Sons.

[9]. Zhang, Y. (2020). Stock Market Price Movement Prediction Using LSTM Deep Learning Model. In International Conference on Advanced Data Mining and Applications (pp. 275-285). Springer, Cham.

[10]. Zuo, Y., & Wang, H. (2020). Deep Learning for Time Series Forecasting: A Survey. arXiv preprint arXiv:2004.13408.

[11]. Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules andthe stochastic properties of stock returns. The Journal of Finance, 47, 1731–1764.

[12]. Cochrane, J. H. (2007). The dog that did not bark: A defense of returns predictability. Review of Financial Studies, 21, 1533–1575.

[13]. Daubechies, I. (1992). Ten lectures on wavelets. Philadelphia, PA: SIAM(Societyfor Industrial and Applied Mathematics).

[14]. Ferreira, M. I., & Santa-Clara, P. (2011). Forecasting stock market returns: The sumof the parts is more than the whole. Journal of Financial Economics, 100, 514–537.

[15] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020, doi: 10.1007/s12652-020-01839-w.

[16] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," J. Supercomput., vol. 76, no. 3, pp. 2098–2118, 2020, doi: 10.1007/s11227-017-2228-y.

[17] I. K. Nti, A. Felix Adekoya, · Benjamin, and A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artif. Intell. Rev., vol. 53, pp. 3007–3057, 123AD, doi: 10.1007/s10462-019-09754-z.

[18] J. L. Wang and S. H. Chan, "Stock market trading rule discovery using pattern recognition and technical analysis," Expert Syst. Appl., vol. 33, no. 2, pp. 304–315, Aug. 2007, doi: 10.1016/J.ESWA.2006.05.002.

[19] Hu Z, Zhu J, Tse K. Stocks market prediction using support vector machine. In 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering 2013 Nov 23 (Vol. 2, pp. 115-118). IEEE.