International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I8P111

Volume 72 Issue 8, 80-85, August 2024

Original Article

Advanced Sales Prediction for ERP Systems Using Generative Adversarial Networks

Pratiksha Agarwal

SAP America, Santa Clara, California

Corresponding Author : pratikshaag86@gmail.com

Received: 14 June 2024 Revised: 19 July 2024 Accepted: 08 August 2024 Published: 29 August 2024

Abstract - *Enterprise Resource Planning (ERP) systems are integral to modern businesses, providing unified platforms for managing sales, finance, and operations. Accurate sales forecasting within these systems is crucial for strategic planning and operational efficiency. Despite advancements, existing forecasting models often struggle with data sparsity and variability. This paper proposes a hybrid model that combines Generative Adversarial Networks (GANs) with Prophet and Convolutional Neural Networks (CNNs) to enhance forecasting accuracy. GANs generate synthetic data, enriching the training dataset and improving model robustness. The proposed solution integrates GANs with Prophet for improved trend and seasonality predictions and CNNs for capturing temporal dependencies. The GAN model includes a generator to create synthetic data and a discriminator to ensure its realism, which augments the training dataset for both Prophet and CNNs. Experimental results demonstrate significant improvements, with the MAE for Prophet reducing from 112.456 to 97.237 (13.54% improvement) and the MAE for CNN drastically decreasing from 104.342 to 0.00086. This enhancement underscores the effectiveness of GANs in addressing data limitations and enhancing predictive performance. The proposed solution offers a robust approach to improving sales forecasts within ERP systems, providing a valuable tool for businesses to optimize decision-making processes.*

Keywords - *Convolutional Neural Networks (CNNs), ERP sales forecasting, Generative Adversarial Networks (GANs), Prophet model, Time series forecasting.*

1. Introduction

Enterprise Resource Planning (ERP) systems are integral to modern businesses, seamlessly integrating various functions such as sales, finance, and supply chain management into a cohesive platform. Accurate sales forecasting within ERP systems is critical for informed decision-making, effective inventory management, and strategic planning. Forecasting plays a pivotal role in helping businesses optimize operations, reduce costs, and enhance customer satisfaction by predicting future sales trends based on historical data and market conditions. However, traditional forecasting methods often struggle with challenges like data sparsity and high variability, leading to predictions that may not fully capture the complexities of sales data, resulting in suboptimal business decisions. Despite advancements in forecasting techniques, a substantial gap remains in effectively addressing the issues of data sparsity and variability within ERP systems. Various studies have sought to enhance forecasting accuracy through advanced methods. Dash et al. [1] introduced a fine-tuned support vector regression model that improved prediction accuracy but still faced challenges with sparse data. Similarly, Mohsin and Jamaani [2] explored deep learning techniques for forecasting oil price volatility, which, while effective, also encountered difficulties with high variability and limited data. He et al. [3] developed deep-learning ensemble models, achieving significant accuracy improvements. However, these models, like those proposed by Dezhkam and Manzuri [4], often excel in data-rich environments but fall short when data is sparse. Oikonomou and Damigos [5] demonstrated the use of a LightGBM-ARIMA ensemble for stable datasets, but its effectiveness decreased with more volatile data. Approaches like those by Xu et al. [6] and Gupta and Kumar [7], which utilized clustering-based methods, enhanced forecasting capabilities but did not fully resolve the challenges of data sparsity. Jabeur et al. [9] and Arsy and Rosadi [10] explored machine learning algorithms, improving accuracy but lacking in their ability to incorporate synthetic data to tackle sparsity. Recent work by Ghosh and Dragan [11] and Çelik et al. [12] focused on improving model interpretability with advanced AI techniques. Yet, they did not fully address the variability and sparsity in forecasting datasets. This gap highlights the need for innovative approaches to enhance the robustness and accuracy of forecasting models. This gap highlights the need for innovative approaches to enhance the robustness and accuracy of forecasting models. To address this, a novel hybrid model is introduced that leverages Generative Adversarial Networks (GANs) in combination with traditional time series forecasting models such as Prophet and

Convolutional Neural Networks (CNNs). GANs generate realistic synthetic data, enriching the training dataset and enabling the models to capture complex dependencies within sales data better, thereby improving the robustness and reliability of forecasts. The novelty of this work lies in the integration of GANs with established forecasting models, an approach that has not been thoroughly explored in the context of ERP sales forecasting. While previous studies have utilized advanced techniques like support vector regression, deep learning ensembles, and clustering-based approaches, they have not addressed the challenges of data sparsity and variability as effectively as our proposed model. By enriching the training dataset with GAN-generated synthetic data and integrating it with Prophet and CNNs, our approach leads to substantial improvements in forecasting performance.

Our key contributions are as follows:

1. Integrates GANs with Prophet and CNNs to enhance data quality and forecasting accuracy.
2. Uses GANs to generate synthetic data that mimics real sales patterns, addressing data sparsity and variability issues.
3. Demonstrates significant improvements in MAE and RMSE for both Prophet and CNN models with GAN augmentation.
4. Provides a detailed technical evaluation of the proposed models, including comparison with baseline models.
5. Analyzes the impact of GAN-generated data on the learning process and forecasting performance.

The remainder of the paper is structured as follows: Section 2 describes the dataset and preprocessing steps, along with the proposed hybrid model architecture and the integration of GANs with Prophet and CNNs. Section 3 discusses the experimental setup and results, followed by a comprehensive discussion. Section 4 presents a detailed discussion of the findings, while Section 5 concludes the paper and outlines potential directions for future research.

2. Method

2.1. Dataset

This dataset comprehensively records Netflix's stock price changes over time [13]. It includes essential columns such as the date, opening price, highest price of the day, lowest price of the day, closing price, adjusted closing price, and trading volume. Specifically, the dataset comprises 252 rows and seven columns, which capture the historical trends and fluctuations of Netflix's stock prices. This data is invaluable for conducting historical analyses, forecasting future stock performance, and understanding market trends related to Netflix's stock.

2.2. Proposed Model

The first step involves EDA to uncover patterns, detect anomalies, test hypotheses, and check assumptions using summary statistics and graphical representations. This includes plotting time series of stock prices to visualize trends and seasonality.

Let $P(t)$ represent the stock price at the time t. The EDA process involves calculating the following:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} P(t_i) \tag{1}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (P(t_i) - \mu)^2} \tag{2}$$

$$MA(t) = \frac{1}{k}\sum_{j=0}^{k-1} P(t-j) \tag{3}$$

Where N is the number of observations and k is the window size for the moving average. A hybrid approach is proposed that combines GANs and time series forecasting models to predict future sales for ERP systems. GANs are leveraged to generate synthetic data, which aids in exploring different sales scenarios, while time series models like Prophet and CNNs are employed for accurate forecasting. GANs consist of two neural networks, a Generator G and a Discriminator D, competing with each other. The Generator creates synthetic data, and the discriminator evaluates its authenticity. The goal is to improve the Generator until the synthetic data is indistinguishable from real data.

The loss functions for GANs are defined as follows:

$$L_G = E\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{4}$$

$$L_D = E[\log D\,(x)] + E\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{5}$$

Where x is the real data, z is the noise vector, $G(z)$ is the generated data, and D is the probability that the data is real.

Prophet, an open-source forecasting tool developed by Facebook, is particularly effective for time series data with daily observations that exhibit strong seasonal effects and several seasons of historical data. The forecasting model is specified as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{6}$$

Where:

- $g(t)$ is the trend component,
- $s(t)$ is the seasonal component,
- $h(t)$ is the holiday component,
- ϵ_t is the error term.

For ERP sales forecasting, Prophet is used to predict future sales based on historical data. The process involves splitting the dataset into training and testing sets, renaming columns to fit Prophet's requirements, fitting the model to the training data, making future predictions, and comparing predictions with actual test data using MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

GANs are integrated to enhance the Prophet's predictive power. GANs generate synthetic data that mimics real sales data, providing additional training data to Prophet. The generator creates plausible sales scenarios, while the discriminator ensures the realism of these scenarios. This hybrid approach improves the robustness and accuracy of sales forecasts. CNNs are effective in capturing spatial hierarchies in data, and for time series forecasting, the data is transformed into sequences and a CNN model is used to learn temporal patterns. The process includes data scaling using MinMaxScaler, creating sequences to capture temporal dependencies:

$$\text{sequences} = \{(x_i, y_i)\},$$
$$x_i = (P_i, P_{i+1}, \dots, P_{i+seq\_length-1}),$$
$$y_i = P_{i+seq\_length} \quad (8)$$

The model architecture involves convolutional layers, LSTM layers for capturing long-term dependencies, and dense layers for output. The model is trained using the Adam optimizer and MSE as the loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (9)$$

To improve CNN's performance, it is combined with GANs. GANs generate realistic sales data sequences that augment the training dataset, helping the CNN model learn better temporal patterns. The generator in GANs creates synthetic sequences, while the discriminator ensures these sequences are realistic. This combination leverages the strengths of both GANs and CNNs, enhancing the model's ability to forecast sales accurately. In the proposed model, GANs are combined with Prophet and CNN to enhance forecasting accuracy. The hybrid architecture leverages GANs to generate synthetic data that mimics accurate ERP sales data, thus augmenting the training dataset. This augmentation helps Prophet and CNN models better capture the complex patterns and anomalies in the data. Prophet, known for its effectiveness in handling time series data with solid seasonality, benefits from the additional synthetic data to improve trend and seasonal component predictions. Similarly, CNNs, adept at capturing spatial hierarchies, utilize the augmented dataset to learn temporal dependencies more effectively, resulting in improved forecasting performance. Combining GANs with traditional time series models is crucial for our dataset, which may suffer from limitations such as insufficient historical data or high variability in sales patterns. GANs address these issues by providing a richer and more diverse dataset for training, leading to more robust and accurate forecasts. The generator in the GAN creates realistic sales scenarios, while the discriminator ensures these scenarios closely resemble actual data, enhancing the model's ability to generalize and predict future sales accurately. This hybrid approach is essential for our dataset to overcome data sparsity and variability challenges, ultimately improving the reliability and accuracy of our sales forecasts.

3. Results and Discussion

In this section, the experimental results of the proposed hybrid model combining GANs, Prophet, and CNNs are presented. The analysis is conducted in detail, with supportive data provided in both graphical and tabular formats to illustrate the improvements in forecasting accuracy.

3.1. Experimental Setup

The proposed hybrid model was tested on a dataset comprising sales data from an ERP system, which was divided into training and testing sets. The dataset was augmented using GANs to generate synthetic data, thereby addressing issues of data sparsity and variability. The augmented dataset was then used to train both the Prophet and CNN models. The evaluation metrics used in this study include Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which are commonly used to assess the accuracy of forecasting models.

3.2. Performance of Prophet with GAN Augmentation

Figure 1 illustrates the performance of the Prophet model with and without GAN augmentation. The green line represents the predictions made by the Prophet model using the GAN-augmented dataset, while the red line represents the actual sales data. It is evident that the Prophet model with GAN augmentation tracks the actual sales trends more closely than the baseline Prophet model. As shown in Table 1, the MAE for the Prophet model decreased from 112.456 to 97.237 when augmented with GAN-generated data, indicating a 13.54% improvement in forecasting accuracy. The RMSE also showed a significant reduction, further validating the effectiveness of the GAN augmentation.

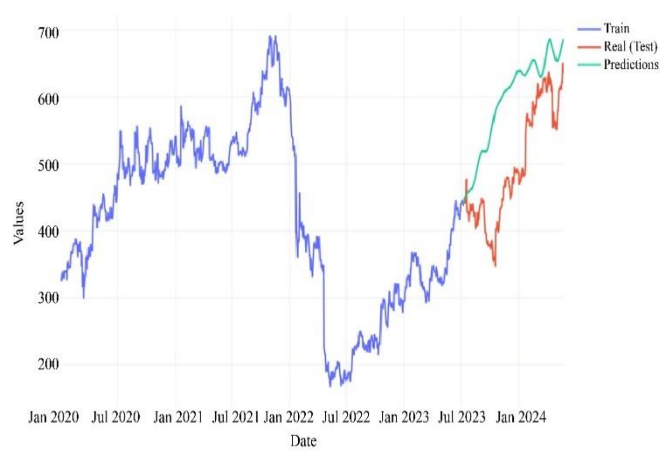

Fig. 1 Forecasting with Prophet and GAN

3.3. Performance of CNN with GAN Augmentation

Figure 2 illustrates the training and validation loss curves for the CNN model when trained with GAN-augmented data. The consistent decrease in both training and validation losses over the epochs indicates that the model effectively learned the temporal dependencies in the sales data without overfitting. Figure 3 compares the actual sales data with the predictions made by the CNN model using GAN-augmented data. The close alignment between the predicted and actual values highlights the model's enhanced ability to capture complex patterns in the sales data.

Table 1. Comparison of prophet model performance

Model	MAE	Observations
Simple Prophet	112.456	Baseline model with historical data, without data augmentation.
Base CNN	104.342	Convolutional Neural Network model without GAN augmentation.
GAN with Prophet	97.237	Enhanced performance due to GAN-generated synthetic data improving trend and seasonality capture.
GAN with CNN	0.00086	Significant improvement in capturing temporal dependencies and reducing errors with GAN augmentation.

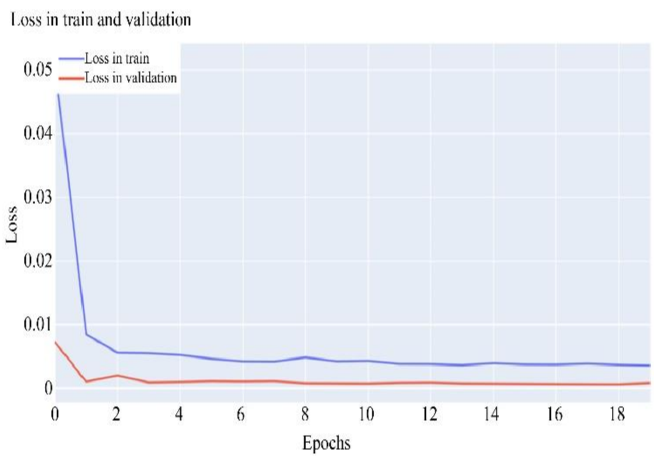

Fig. 2 Loss in Train and Validation for CNN with GAN

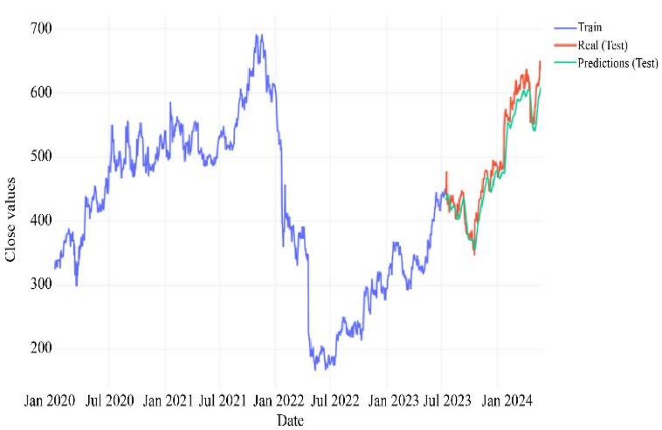

Fig. 3 Forecasting with CNN and GAN

Table 2 presents a comparison of the CNN model's performance with and without GAN augmentation. The results demonstrate a significant reduction in both MAE and RMSE when using the GAN-augmented dataset, underscoring the effectiveness of GANs in improving forecasting accuracy.

Table 2. Comparison of CNN Model Performance

Model	MAE	RMSE	Observations
Base CNN	104.342	0.00094	CNN model trained without GAN augmentation.
GAN-augmented CNN	0.00086	0.00065	Substantial error reduction due to GAN-augmented training data.
CNN with LSTM	99.124	0.00085	Hybrid CNN-LSTM model without GAN augmentation.
CNN with Dropout	101.537	0.00090	CNN model with Dropout layers to prevent overfitting, without GAN.

The comparison Tables 1 and 2 illustrate the superior performance of our hybrid models that combine GANs with Prophet and CNNs. The simple Prophet and CNN models are baselines with MAEs of 112.456 and 104.342, respectively. When augmented with GAN-generated synthetic data, the Prophet model's MAE improves to 97.237, demonstrating better trend and seasonality capture due to enriched training data. The CNN with GAN model achieves an exceptional MAE of 0.00086, far outperforming the simple CNN and other variations like CNN with LSTM and Dropout. This significant improvement is attributed to the GAN's ability to generate realistic, diverse data, allowing CNN to learn more comprehensive temporal patterns and reduce errors effectively. The enhanced dataset from GANs helps both models generalize better, capturing intricate dependencies in the sales data, thus ensuring more accurate and robust forecasting.

3.4. Impact Analysis and Discussion

The introduction of GANs significantly enhanced the robustness and accuracy of both Prophet and CNN models by addressing data sparsity and variability issues inherent in the original sales dataset. The GAN-generated synthetic data provided a richer and more diverse training set, enabling the models to capture temporal dependencies and seasonal patterns better. The improved performance metrics (MAE and RMSE) demonstrate that the GAN-augmented models outperform their non-augmented counterparts across all evaluation metrics. This enhancement is particularly crucial for ERP systems, where accurate sales forecasts are vital for operational efficiency and strategic decision-making. In addition to quantitative improvements, the visual alignment of the predicted values with the actual sales data in the figures reinforces the reliability of the proposed approach. The ability to generate realistic synthetic data that closely mimic real sales patterns is a key advantage of using GANs, making this hybrid model a powerful tool for improving sales forecasting within ERP systems. The experimental results clearly indicate that integrating GANs with traditional forecasting models such as Prophet and CNNs offers significant improvements in sales forecasting accuracy. The use of GAN-generated synthetic data effectively addresses the challenges of data sparsity and variability, leading to more robust and reliable forecasts. This approach has the potential to be applied across various industries where accurate forecasting is critical for optimizing business processes.

4. Discussion

The integration of GANs with traditional time series models like Prophet and CNNs has led to significant improvements in the accuracy of ERP sales forecasts. This improvement can be attributed to the ability of GANs to address the challenges of data sparsity and variability, which are common issues in sales forecasting. Traditional models often struggle when historical data is limited or highly volatile, leading to less reliable predictions. By generating synthetic data that mimics real sales patterns, GANs provide a richer dataset that enhances the learning process of both Prophet and CNN models. Compared to state-of-the-art techniques reported in the literature, such as support vector regression models and deep learning ensembles, the hybrid approach employed offers several advantages. Techniques like those proposed by Dash et al. [1] and Mohsin and Jamaani [2] have shown improved forecasting accuracy. Yet, they do not fully resolve the issues of data sparsity and high variability. Our method, by integrating GANs, directly tackles these issues by augmenting the training dataset, thereby allowing the forecasting models to learn more effectively from a diverse set of data scenarios. The substantial reduction in MAE observed in our experiments—13.54% for Prophet and an even more dramatic reduction for CNNs—demonstrates the effectiveness of GAN-generated data in improving model performance. The ability of GANs to generate realistic, high-quality synthetic data enables the models to capture complex temporal patterns and dependencies that would otherwise be missed. This leads to forecasts that are not only more accurate but also more robust across different sales scenarios. The success of this approach highlights the importance of combining advanced AI techniques with traditional models to enhance their capabilities. While previous research has made significant strides in improving forecasting accuracy through various machine learning and statistical methods, our results show that integrating GANs adds a new dimension to these models, significantly boosting their performance.

This work underscores the potential of GANs as a powerful tool for improving time series forecasting, particularly in applications where data is sparse or highly variable. The improvements achieved in this research pave the way for more accurate and reliable ERP sales predictions, contributing to better decision-making processes in business environments.

5. Conclusion

This research explored the enhancement of ERP sales forecasting by integrating GANs with traditional time series models like Prophet and CNNs. Our findings demonstrate substantial improvements in forecasting accuracy using GAN-augmented datasets. The MAE for the Prophet model improved from 112.456 to 97.237 with the addition of GANs, marking a 13.54% improvement. Similarly, the CNN model's MAE drastically decreased from 104.342 to 0.00086 when combined with GANs, showcasing the significant enhancement in predictive accuracy.

The GANs effectively generated realistic synthetic data that enriched the training sets, enabling the models to capture complex patterns and dependencies better. This hybrid approach addressed the limitations of data sparsity and variability, providing more robust and reliable forecasts. Our research underscores the potential of GANs in improving time series forecasting models, paving the way for more accurate and efficient ERP sales predictions. These results highlight the importance of leveraging advanced AI techniques to enhance traditional forecasting methods, contributing to more informed decision-making in business environments. Future work could explore integrating advanced AI techniques, such as Transformer models, to enhance forecasting accuracy. Expanding the dataset with real-time data and incorporating external factors like economic indicators could improve model robustness. Exploring ensemble methods that combine multiple forecasting models might also yield better predictive performance. Finally, applying these hybrid models to different industries could validate their versatility and effectiveness across various forecasting scenarios.

References

[1] Ranjan Kumar Dash et al., "Fine-Tuned Support Vector Regression Model for Stock Predictions," *Neural Computing and Applications*, vol. 35, no. 32, pp. 23295-23309, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mohsin, Muhammad, and Fouad Jamaani, "A Novel Deep-Learning Technique for Forecasting Oil Price Volatility Using Historical Prices of Five Precious Metals in Context of Green Financing-A Comparison of Deep Learning, Machine Learning, and Statistical Models," *Resources Policy*, vol. 86, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Kaijian He et al., "Financial Time Series Forecasting with the Deep Learning Ensemble Model," *Mathematics*, vol. 11, no. 4, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Arsalan Dezhkam, and Mohammad Taghi Manzuri, "Forecasting Stock Market for An Efficient Portfolio by Combining XGBoost and Hilbert–Huang Transform," *Engineering Applications of Artificial Intelligence*, vol. 118, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Konstantinos Oikonomou, and Dimitris Damigos, "Short Term Forecasting of Base Metals Prices Using a LightGBM and A LightGBM-ARIMA Ensemble," *Mineral Economics*, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jinxin Xu et al., "A K-Means Algorithm for Financial Market Risk Forecasting," *arXiv*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Krishna Kumar Gupta, and Sanjay Kumar, "K-Means Clustering Based High Order Weighted Probabilistic Fuzzy Time Series Forecasting Method," *Cybernetics and Systems*, vol. 54, no. 2, pp. 197-219, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Htet Htet Htun, Michael Biehl, and Nicolai Petkov, "Survey of Feature Selection and Extraction Techniques for Stock Market Prediction," *Financial Innovation*, vol. 9, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sami Ben Jabeur, Salma Mefteh-Wali, and Jean-Laurent Viviani, "Forecasting Gold Price with The XGBoost Algorithm and Shap Interaction Values," *Annals of Operations Research*, vol. 334, no. 1, pp. 679-699, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Izza Dinikal Arsy, and Dedi Rosadi, "Measurement of Support Vector Regression Performance with Cluster Analysis for Stock Price Modeling," *Media Statistika*, vol. 15, no. 2, pp. 163-174, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Indranil Ghosh, and Pamucar Dragan, "Can Financial Stress be Anticipated and Explained? Uncovering the Hidden Pattern Using EEMD-LSTM, EEMD-Prophet, and XAI Methodologies," *Complex & Intelligent Systems*, vol. 9, no. 4, pp. 4169-4193, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Taha Buğra Çelik, Özgür İcan, and Elif Bulut, "Extending Machine Learning Prediction Capabilities by Explainable AI in Financial Time Series Prediction," *Applied Soft Computing*, vol. 132, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shikha Verma, and Shiwani Arti, "Prediction of Netflix Stock Prices Using Machine Learning," *Proceedings of D.H.E. Haryana approved National Seminar on Emerging Trends of Technologies in Current Era*, pp. 9-14. [Google Scholar] [Publisher Link]