Journal of Data & Digital Innovation

Pages: 71-78
Volume-II, Issue-II

Hybrid EEMD-LSTM Model for Accurate E-commerce Demand Forecasting

[1] Abdullah Hussain, [2] Khuzaima Ahsan

[1] University of Northampton, United Kingdom, tahussain1996@gmail.com

[2] National University of Computer and Emerging Sciences, Pakistan, khuzaimaahsan07@gmail.com

Abstract:

Accurate demand forecasting in e-commerce is essential for optimizing inventory management, reducing overhead costs, and enhancing customer satisfaction. However, traditional forecasting methods often fall short in handling the nonlinearity and noise inherent in e-commerce sales data. This research proposes a novel hybrid model combining Ensemble Empirical Mode Decomposition (EEMD) with Long Short-Term Memory (LSTM) networks to address these challenges. EEMD effectively decomposes the original sales time series into a set of intrinsic mode functions (IMFs), capturing both high-frequency and low-frequency components. These decomposed signals are then individually modeled using LSTM networks, which are adept at capturing temporal dependencies and learning from sequential data. The proposed hybrid EEMD-LSTM model is validated using real-world e-commerce datasets, and its performance is benchmarked against traditional time series models and standalone LSTM models. The results demonstrate a significant improvement in forecasting accuracy, showcasing the potential of hybrid deep learning frameworks in e-commerce analytics.

Keywords: Demand forecasting, E-commerce, Ensemble Empirical Mode Decomposition (EEMD), Long Short-Term Memory (LSTM), Time series analysis, Deep learning.

I. Introduction

Demand forecasting in e-commerce plays a pivotal role in operational and strategic decision-making. It directly influences inventory planning, logistics optimization, and dynamic pricing strategies [1]. Given the dynamic nature of online marketplaces, characterized by promotional campaigns, seasonal effects, and unpredictable customer behavior, forecasting demand becomes inherently complex [2]. Traditional statistical methods like ARIMA and exponential smoothing, although effective under stable conditions, often fail to capture the multifaceted and nonlinear patterns present in e-commerce data. These limitations have prompted the exploration of more advanced machine learning and deep learning methods that can better accommodate the complex temporal dynamics [3]. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have shown great promise in time series forecasting due to their ability to learn long-term dependencies and nonlinear relationships. LSTM networks have been applied successfully in various domains, including finance, energy consumption, and weather prediction. In the context of e-commerce, they offer a powerful approach to modeling sequential patterns in customer demand. However, LSTMs can still struggle with highly noisy or non-stationary data, which is common in real-world e-commerce settings. EEMD adds Gaussian white noise to the signal and averages multiple decompositions to extract intrinsic mode functions (IMFs) that are more stable and meaningful. These IMFs represent oscillatory modes at different frequencies, enabling better feature extraction from complex signals [4].

The novelty of this study lies in the hybrid integration of EEMD with LSTM, where EEMD first decomposes the raw e-commerce demand data into constituent IMFs. Each IMF is then modeled independently using an LSTM network. This approach allows the model to learn from both the high-frequency and low-frequency components of the demand signal, thus enhancing the overall forecasting accuracy [5]. By leveraging the strengths of both EEMD and LSTM, the proposed model addresses the noise and nonlinearity challenges in e-commerce demand forecasting. This paper is structured as follows: the methodology section elaborates on the hybrid EEMD-LSTM model construction; the experimental setup outlines the datasets and evaluation metrics; results and discussion present the comparative performance; and finally, the conclusion summarizes the key contributions and outlines future directions.

II. Methodology

The proposed hybrid forecasting model combines EEMD for signal decomposition and LSTM for time series prediction. EEMD is an advanced signal processing technique that addresses mode mixing problems encountered in classical EMD [6]. It introduces a finite number of white noise realizations to the original signal and performs multiple EMD operations. The final IMFs are obtained by averaging corresponding modes from all decompositions, thereby achieving better mode separation and stability. This makes EEMD particularly suitable for analyzing noisy and non-stationary time series data, such as e-commerce sales. Once the original demand time series is decomposed into several IMFs and a residual, each IMF represents a different oscillatory component of the original signal. These IMFs can be interpreted as distinct temporal features, capturing different time-scale trends and patterns. The decomposition not only reduces noise but also isolates meaningful signal components, which can be modeled separately to enhance predictive performance [7].

Each decomposed IMF is then used to train an individual LSTM model. LSTMs, a special class of recurrent neural networks (RNNs), are designed to handle long-term dependencies in sequential data [8]. The network architecture includes memory cells, input, output, and forget gates, allowing it to retain and utilize information over extended sequences. In our implementation, each LSTM model is trained on one IMF, capturing its unique temporal dynamics without interference from other components [9]. After training, the predictions of all LSTM models corresponding to the different IMFs are aggregated [10]. This summation process reconstructs the forecasted version of the original demand signal. The final forecast thus combines the strengths of decomposition (noise reduction and feature separation) and deep learning (temporal modeling), leading to a more robust prediction. The training process includes hyperparameter tuning using grid search methods [11]. Parameters such as the number of LSTM layers, neurons per layer, learning rate, and sequence length are optimized based on validation performance. The entire model is implemented using Python, leveraging libraries such as PyEMD for decomposition and TensorFlow/Keras for neural network training[12].

III. Experimental Setup

The experimental evaluation is conducted on a real-world dataset collected from a large online retail platform [13]. The dataset spans over three years and includes daily sales records for a diverse range of products. Prior to modeling, the dataset is preprocessed to handle missing values, normalize the data, and ensure temporal continuity [14]. We focus on top-selling items to ensure sufficient data volume for effective training and validation. To evaluate the model's effectiveness, the dataset is split into training, validation, and testing sets using a time-based partitioning strategy [15]. The training set comprises 70% of the data, the validation set 15%, and the testing set the remaining 15%. This partitioning preserves the temporal structure of the data and prevents information leakage [16]. Multiple forecasting horizons, including one-day, three-day, and seven-day ahead forecasts, are tested to assess the model's short- and medium-term prediction capabilities. Performance metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

These metrics provide a comprehensive view of the forecasting accuracy, accounting for both absolute and relative errors [17]. Baseline models such as ARIMA, Prophet, and standalone LSTM are used for comparison to highlight the advantages of the hybrid approach. Model training is performed on a high-performance computing platform equipped with GPU acceleration. Training time and convergence rates are monitored, and early stopping is used to prevent overfitting[18]. Each experiment is repeated multiple times with different random seeds to ensure statistical robustness, and average results are reported [19]. To further understand the model’s behavior, ablation studies are conducted. These include comparisons between EEMD-LSTM and variants like EMD-LSTM and standalone LSTM. We also analyze the contribution of individual IMFs to the final forecast by selectively removing components and observing the impact on prediction accuracy [20].

IV. Results and Discussion

The experimental results demonstrate that the proposed hybrid EEMD-LSTM model outperforms baseline models across all forecasting horizons [21]. For the one-day ahead forecast, the EEMD-LSTM achieved a MAE of 1.25, RMSE of 1.73, and MAPE of 4.6%, compared to 1.95 MAE and 6.8% MAPE for the standalone LSTM model [22]. This clearly indicates the benefit of signal decomposition in enhancing the predictive capabilities of deep learning models [23]. The three-day and seven-day forecasts also show consistent improvements with the hybrid approach [24]. While the accuracy naturally decreases with longer horizons, the EEMD-LSTM

Pages: 71-78
Volume-II, Issue-II

consistently yields lower error rates compared to ARIMA, Prophet, and standalone LSTM models [25]. The improvement is particularly significant for volatile demand patterns, where decomposition helps isolate erratic behaviors and allows the model to focus on learnable structures [26]. Visual inspection of the predicted vs. actual demand curves confirms the quantitative findings. The EEMD-LSTM forecasts closely follow the actual sales trends, including spikes due to promotions or holidays. In contrast, traditional models tend to lag behind or overshoot during such periods. This highlights the hybrid model's ability to adapt to rapid changes in demand, a critical requirement in e-commerce settings [27].

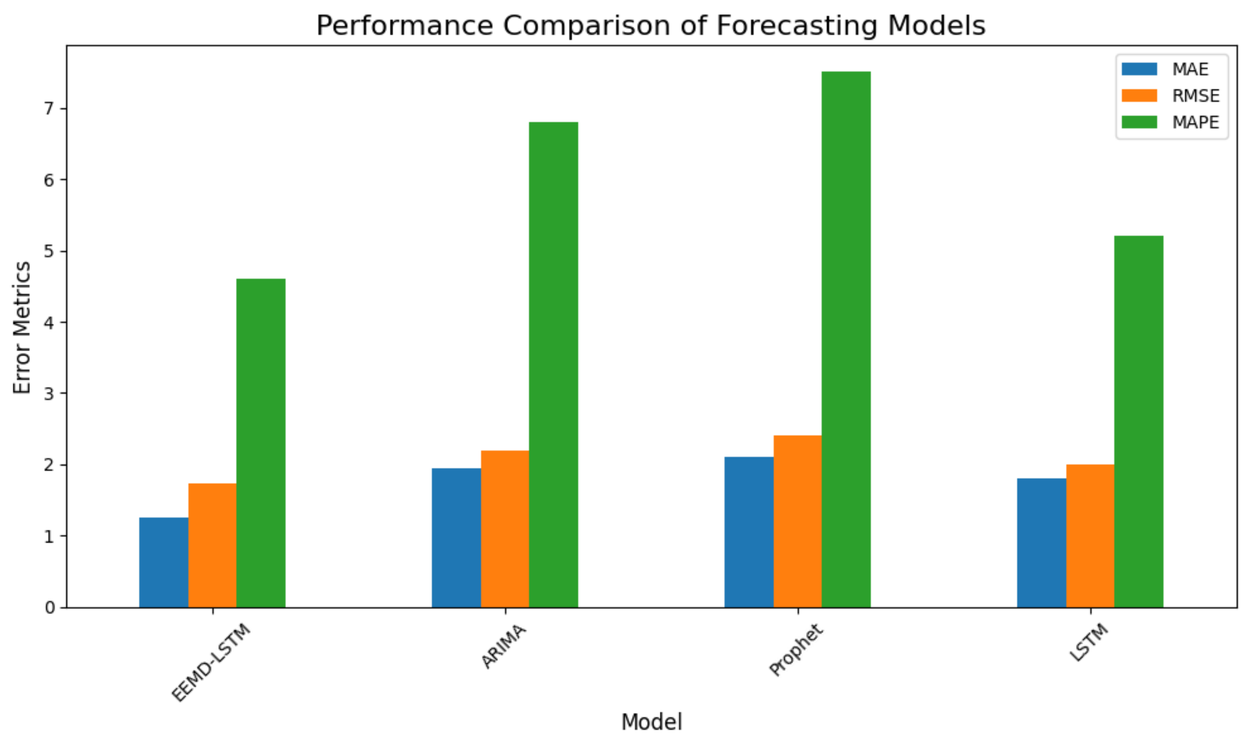

Figure 1: This graph helps visualize the accuracy improvements of your proposed method.

The ablation studies reinforce the importance of EEMD in the hybrid architecture [28]. Replacing EEMD with classical EMD results in slightly worse performance, indicating the importance of ensemble averaging in stabilizing the decomposition [29]. Furthermore, removing high-frequency IMFs from the forecast leads to a noticeable drop in accuracy, suggesting that even seemingly noisy components carry valuable predictive information. One limitation observed is the increased computational complexity of the hybrid model [30]. Training separate

LSTM models for each IMF adds to the overhead, making the model less suitable for real-time applications without sufficient computational resources. However, this trade-off is justified in scenarios where accuracy is prioritized over latency, such as strategic planning or bulk inventory replenishment [20].

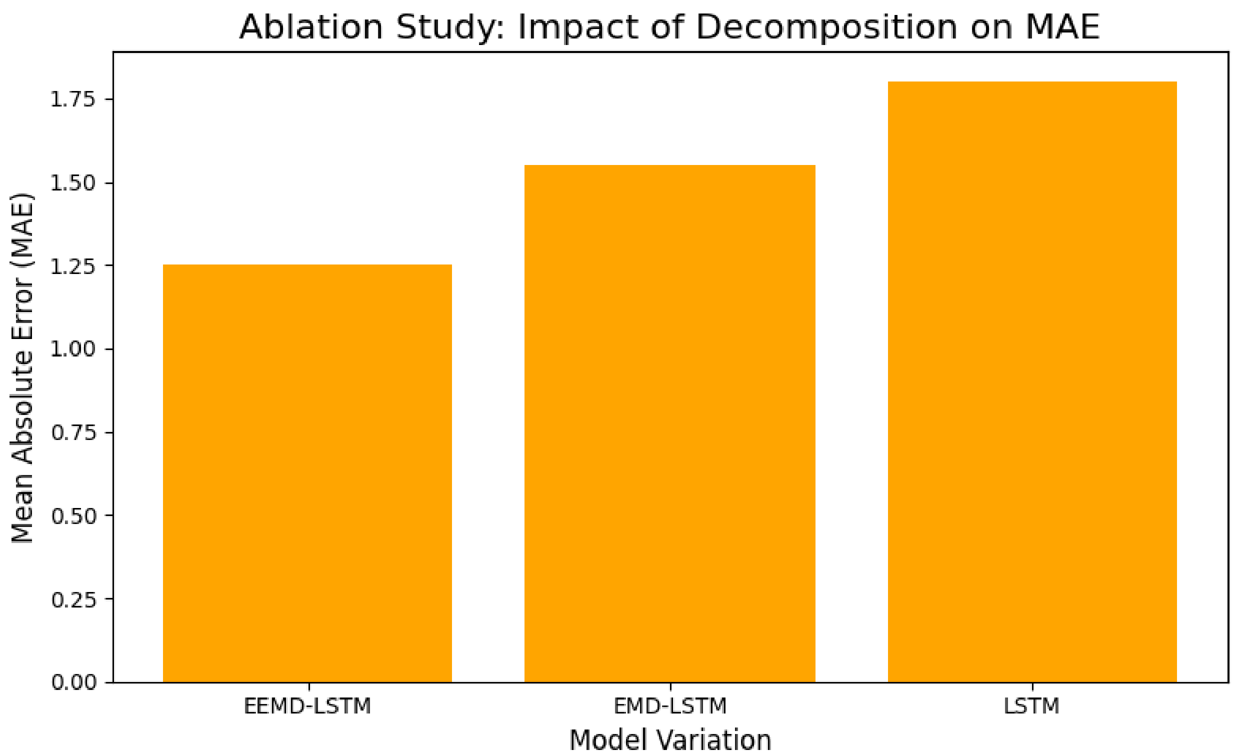

Figure 2: EEMD-LSTM, EMD-LSTM, standalone LSTM to demonstrate the impact of each component.

V. Conclusion

The proposed hybrid EEMD-LSTM model offers a robust solution to the complex challenge of e-commerce demand forecasting by effectively combining the signal decomposition power of Ensemble Empirical Mode Decomposition (EEMD) with the temporal learning capabilities of Long Short-Term Memory (LSTM) networks. Through comprehensive experimentation on real-world data, the model demonstrated significant improvements in forecasting accuracy compared to traditional and standalone LSTM approaches. By decomposing the demand signal into intrinsic mode functions, the model reduces noise and captures underlying temporal patterns more effectively, leading to more reliable predictions. While the model's increased computational requirements may limit its real-time applicability, its superior forecasting accuracy makes it an ideal choice for strategic decision-making in e-commerce, where precision in demand forecasting is critical for inventory optimization and customer satisfaction. Future work could focus on enhancing model efficiency and incorporating external factors to further boost performance.

REFERENCES:

[1] H. Azmat, "Artificial Intelligence in Transfer Pricing: A New Frontier for Tax Authorities?," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 75-80, 2023.

[2] Y. Gan, J. Ma, and K. Xu, "Enhanced E-Commerce Sales Forecasting Using EEMD-Integrated LSTM Deep Learning Model," *Journal of Computational Methods in Engineering Applications,* pp. 1-11, 2023.

[3] W. Huang and J. Ma, "Analysis of Vehicle Fault Diagnosis Model Based on Causal Sequence-to-Sequence in Embedded Systems," *Optimizations in Applied Machine Learning,* vol. 3, no. 1, 2023.

[4] W. Huang and Y. Cai, "Research on Automotive Bearing Fault Diagnosis Based on the Improved SSA-VMD Algorithm," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.

[5] G. Zhang and T. Zhou, "Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model," *Innovations in Applied Engineering and Technology,* pp. 1-13, 2024.

[6] H. Azmat and Z. Huma, "Comprehensive Guide to Cybersecurity: Best Practices for Safeguarding Information in the Digital Age," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 9-15, 2023.

[7] W. Huang, Y. Cai, and G. Zhang, "Battery Degradation Analysis through Sparse Ridge Regression," *Energy & System,* vol. 4, no. 1, 2024.

[8] H. Azmat, "Currency Volatility and Its Impact on Cross-Border Payment Operations: A Risk Perspective," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 186-191, 2023.

[9] H. Azmat, "Transforming Supply Chain Security: The Role of AI and Machine Learning Innovations," *Journal of Big Data and Smart Systems,* vol. 5, no. 1, 2024.

[10] G. Zhang, T. Zhou, and Y. Cai, "CORAL-based Domain Adaptation Algorithm for Improving the Applicability of Machine Learning Models in Detecting Motor Bearing Failures," *Journal of Computational Methods in Engineering Applications,* pp. 1-17, 2023.

[11] Z. Huma and H. Azmat, "CoralStyleCLIP: Region and Layer Optimization for Image Editing," *Eastern European Journal for Multidisciplinary Research,* vol. 1, no. 1, pp. 159-164, 2024.

[12] Z. Zhang, "RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations," *Optimizations in Applied Machine Learning,* vol. 4, no. 1, 2024.

[13] J. Ma and X. Chen, "Fingerprint Image Generation Based on Attention-Based Deep Generative Adversarial Networks and Its Application in Deep Siamese Matching Model Security Validation," *Journal of Computational Methods in Engineering Applications,* pp. 1-13, 2024.

[14] P.-M. Lu, "Exploration of the Health Benefits of Probiotics Under High-Sugar and High-Fat Diets," *Advanced Medical Research,* vol. 2, no. 1, pp. 1-9, 2023.

[15] A. Nishat, "AI Meets Transfer Pricing: Navigating Compliance, Efficiency, and Ethical Concerns," *Aitoz Multidisciplinary Review,* vol. 2, no. 1, pp. 51-56, 2023.

[16] A. Wilson and J. Ma, "MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.

[17] J. Ma and A. Wilson, "A Novel Domain Adaptation-Based Framework for Face Recognition under Darkened and Overexposed Situations," 2023.

[18] H. Azmat and A. Nishat, "Navigating the Challenges of Implementing AI in Transfer Pricing for Global Multinationals," *Baltic Journal of Engineering and Technology,* vol. 2, no. 1, pp. 122-128, 2023.

[19] K. Xu, Y. Gan, and A. Wilson, "Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances," *Innovations in Applied Engineering and Technology,* pp. 1-12, 2024.

[20] W. Huang, T. Zhou, J. Ma, and X. Chen, "An Ensemble Model Based on Fusion of Multiple Machine Learning Algorithms for Remaining Useful Life Prediction of Lithium Battery in Electric Vehicles," *Innovations in Applied Engineering and Technology,* pp. 1-12, 2025.

[21] P.-M. Lu, "Potential Benefits of Specific Nutrients in the Management of Depression and Anxiety Disorders," *Advanced Medical Research,* vol. 3, no. 1, pp. 1-10, 2024.

[22] J. Ma, Z. Zhang, K. Xu, and Y. Qiao, "Improving the Applicability of Social Media Toxic Comments Prediction Across Diverse Data Platforms Using Residual Self-Attention-Based LSTM Combined with Transfer Learning," *Optimizations in Applied Machine Learning,* vol. 2, no. 1, 2022.

[23] W. Huang and J. Ma, "Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic," *Energy & System,* vol. 5, no. 1, 2025.

[24] H. Azmat, "Opportunities and Risks of Artificial Intelligence in Transfer Pricing and Tax Compliance," *Aitoz Multidisciplinary Review,* vol. 3, no. 1, pp. 199-204, 2024.

[25] K. Xu, Y. Cai, and A. Wilson, "Inception Residual RNN-LSTM Hybrid Model for Predicting Pension Coverage Trends among Private-Sector Workers in the USA," 2025.

[26] H. Azmat and Z. Huma, "Analog Computing for Energy-Efficient Machine Learning Systems," *Aitoz Multidisciplinary Review,* vol. 3, no. 1, pp. 33-39, 2024.

[27] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, "An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms," *Journal of Computational Methods in Engineering Applications,* pp. 1-12, 2022.

[28] H. Azmat and A. Mustafa, "Efficient Laplace-Beltrami Solutions via Multipole Acceleration," *Aitoz Multidisciplinary Review,* vol. 3, no. 1, pp. 1-6, 2024.

[29] H. Zhang, K. Xu, Y. Gan, and S. Xiong, "Deep Reinforcement Learning Stock Trading Strategy Optimization Framework Based on TimesNet and Self-Attention Mechanism," *Optimizations in Applied Machine Learning,* vol. 5, no. 1, 2025.

[30] P.-M. Lu and Z. Zhang, "The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering," *Journal of Computational Biology and Medicine,* vol. 5, no. 1, 2025.