# Sales Forecasting Study Based on a Composite Model of Deep Learning and Random Forest

Ziyao Wang[a], Yining Liu[*]

School of Economics and Management, Xidian University, Xi'an, China

[a]wangziyao0725@163.com, *13309181031@163.com

**Abstract.** In the current rapidly changing economic market environment, accurately forecasting sales is crucial for optimizing enterprise resources and enhancing market competitiveness. Existing prediction models often fail to fully handle complex data relationships and long-term dependencies, which limits the accuracy and practicality of the forecasts. To address these limitations, this study introduces a composite model that integrates deep learning with Random Forest (RF) to significantly enhance predictive performance. This model employs Convolutional Neural Network (CNN) to capture complex features of time-series data and uses Bidirectional Long Short-Term Memory network (BiLSTM) to manage dependencies in data both before and after, while RF reduce overfitting through multiple decision trees and achieve feature fusion through joint training, thereby optimizing prediction accuracy. Experimental results demonstrate that this model outperforms traditional models on all evaluation metrics, particularly showing exceptional adaptability in highly volatile markets with its accuracy and stability.

**Keywords:** Market Economy, Sales Forecasting, Big Data Analysis, Deep Learning, Random Forest.

## 1 Introduction

In today's fast-paced market economy, accurate sales forecasting is crucial for businesses to optimize resources, manage supply chains effectively, and maintain competitive advantages. The complexity and volume of data generated daily pose significant challenges in accurately predicting sales trends. Failure to adequately forecast can result in inventory discrepancies, lost revenue, and reduced market responsiveness [1]. Addressing this challenge is essential for enhancing operational efficiency and market adaptability.

Historically, statistical methods such as ARIMA and exponential smoothing have been the cornerstone of sales forecasting [2]. These techniques have proven effective in stable market conditions with linear patterns. However, their performance degrades with the non-linear and complex data relationships characteristic of modern big data environments. Additionally, traditional models often fail to capture long-term dependencies and rapid market shifts, leading to suboptimal decision-making.



K. Zhang et al. (eds.), *Proceedings of the 5th International Conference on Economic Management and Big Data Application (ICEMBDA 2024)*, Advances in Economics, Business and Management Research 313,
https://doi.org/10.2991/978-94-6463-638-3_20

In recent years, advances in machine learning have introduced more robust forecasting models, such as support vector machines and basic neural networks, which have improved accuracy over traditional statistical methods. Deep learning techniques have shown promise in handling non-linear data and extracting complex patterns. Yet, these models can be computationally intensive and often require large datasets and extensive training times, which might not be feasible for all businesses [3].

This study introduces a composite model that integrates the strengths of deep learning with the robustness of random forests to enhance predictive accuracy and computational efficiency. Firstly, the hybrid architecture that synergistically blends deep learning with ensemble techniques; secondly, the use of CNN and BiLSTM for dynamic feature extraction and dependency modeling, a novel approach in sales forecasting; and thirdly, the incorporation of random forests to enhance prediction stability and reduce the likelihood of overfitting. Experimental results demonstrate that our model outperforms existing methods significantly, showing lower error rates and higher stability in volatile markets.

## 2 Related Work

Research on sales forecasting has traditionally been segmented into several distinct approaches based on the methodology and data handling capabilities. The primary categories include statistical models, machine learning techniques, and hybrid models. Each category has developed substantially over the years, influenced by advancements in computational technology and the increasing availability of data.

Statistical approaches such as ARIMA and Exponential Smoothing have long dominated the field of sales forecasting. A seminal work by Box and Jenkins [4] laid the groundwork for using ARIMA models in time series analysis, which has been widely cited and used for straightforward time series data with linear trends and seasonality. However, these models often fall short in handling non-linear patterns and high volatility, as noted in the comprehensive reviews by Hyndman and Athanasopoulos[5], which question their effectiveness in today's rapidly changing market conditions.

The advent of machine learning has introduced more flexible models capable of capturing complex non-linear relationships within data. Notably, works like those by Ahmed, Atiya, Gayar, and El-Shishiny[6] have demonstrated the superiority of neural networks in handling multivariate and non-linear forecasting problems. Nonetheless, these models often require large datasets and extensive tuning of parameters, which can limit their practical application in smaller or more constrained environments.

Hybrid models that combine statistical and machine learning approaches have gained popularity for their ability to leverage the strengths of both methodologies [7]. A notable example is the work by Pai et al [8], which integrates ARIMA with deep learning techniques to stabilize the predictions in volatile markets. While these models show improved accuracy, they can be complex and computationally expensive to implement, potentially limiting their use to scenarios where high computational resources are available.

Despite these advancements, there remains a gap in effectively integrating the robustness of machine learning models with the speed and simplicity of statistical methods, especially in scenarios with limited data or computational resources. Current literature still struggles with creating models that can quickly adapt to sudden market changes without extensive retraining.

Therefore, this study plans to adopt a novel approach by developing a lightweight, adaptive hybrid model that combines simplified deep learning architectures with ensemble methods to provide both accuracy and efficiency. This model aims to address the identified gap by being capable of rapid adaptation to changing data patterns and providing reliable forecasts even with limited input data. The expected outcome is a versatile forecasting tool that delivers high accuracy while being feasible for real-time applications in various business environments.

## 3 Methodology

The methodology section elucidates the design and implementation of the proposed CNN-BiLSTM-RF hybrid model, aimed at achieving high accuracy in sales forecasting by leveraging the strengths of CNN, BiLSTM, and RF. This model is engineered to address the limitations of existing forecasting approaches by enhancing feature extraction, sequence modeling, and prediction stability.

### 3.1 Convolutional Neural Network (CNN)

The CNN component serves as the primary feature extractor in our model, processing input time-series data to identify local patterns and temporal correlations that are not readily apparent. The CNN architecture is composed of multiple convolutional layers, each consisting of a set of learnable filters that convolve across the input data:

$$h_t^l = f(W^l * h_t^{l-1} + b^l) \tag{1}$$

Here, $h_t^l$ represents the output of layer $l$ at time $t$, $W^l$ and $b^l$ are the weights and biases of the convolutional layer $l$,$*$ denotes the convolution operation, and $f$ is a nonlinear activation function, typically ReLU. The convolutional layers are followed by pooling layers that reduce the dimensionality of the data, helping to make the feature maps more abstract and invariant to small shifts and distortions in the input data[9].

### 3.2 Bidirectional Long Short-Term Memory Network (BiLSTM)

Following feature extraction, the BiLSTM layers are employed to enhance the temporal resolution of the features captured by the CNN. BiLSTM are particularly adept at capturing both past (backward) and future (forward) contexts, a crucial factor for accurate forecasting:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t) \tag{2}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t+1}, x_t) \tag{3}$$

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \tag{4}$$

$\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the hidden states of the forward and backward LSTM at time t, respectively, and $\oplus$ denotes the concatenation of the forward and backward hidden states, which are then passed to the subsequent layer or used for prediction[10].

### 3.3 Random Forest (RF)

The Random Forest algorithm is used as the final layer in our model architecture. It aggregates the features processed by the CNN and BiLSTM layers, providing robust predictions by averaging the outputs of multiple decision trees, each trained on different parts of the feature set and data samples [11]:

$$y_t = \frac{1}{N}\sum_{i=1}^{N} Tree_i\big(concat(h_t)\big) \tag{5}$$

Where $N$ is the number of decision trees, $\text{Tree}_i$ represents the prediction of the $i$-th decision tree, and $\text{concat}(h_t)$ is the concatenated final features from the BiLSTM layer. This approach effectively reduces variance and overfitting, making the model robust against diverse market conditions and noise.

### 3.4 Integration of Component

The integration of CNN, BiLSTM, and RF components is designed to harness the complementary capabilities of these models. The CNN's ability to extract detailed features from raw time-series data, combined with the BiLSTM's capacity to understand time dependencies, and the RF's robustness and generalization, culminates in a powerful forecasting tool. This synergy is achieved through careful tuning of each component and iterative training, ensuring that the features and predictions are well-aligned with the underlying data dynamics.

## 4 Experiment

In the experimental section, this paper utilizes a dataset consisting of two consecutive years of sales data from a particular company to demonstrate the efficacy of a novel composite model that combines deep learning with RF techniques. The experiments validate the model's ability to accurately predict sales trends, leveraging the robustness of RF to reduce overfitting while enhancing predictive stability.

### 4.1 Exploratory Data Analysis (EDA)

The seasonal decomposition of the time series, as shown in Figure 1, reveals factors driving sales fluctuations. Trend components indicate a continuous increase in usage

throughout the year, peaking in summer and autumn, reflecting higher consumer engagement during these seasons. Seasonal factors reveal higher sales in summer compared to winter, suggesting that seasonal marketing strategies may be beneficial. Residuals provide an analysis of irregular or random fluctuations, indicating anomalies that may stem from external factors or data collection errors.

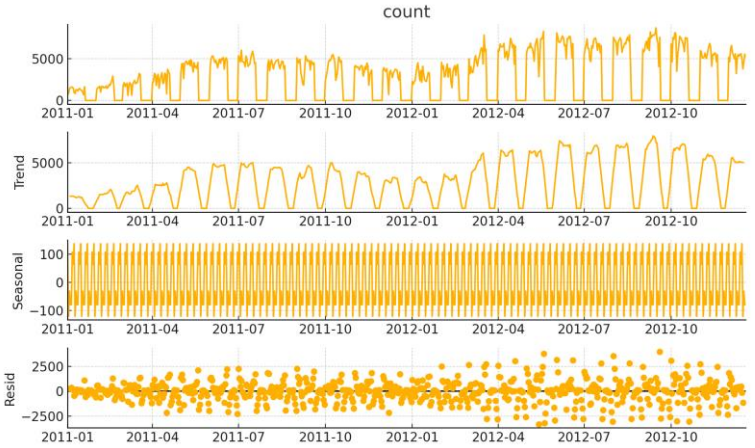


**Fig. 1.** Analysis of Sales Trends, Seasonality, and Residuals

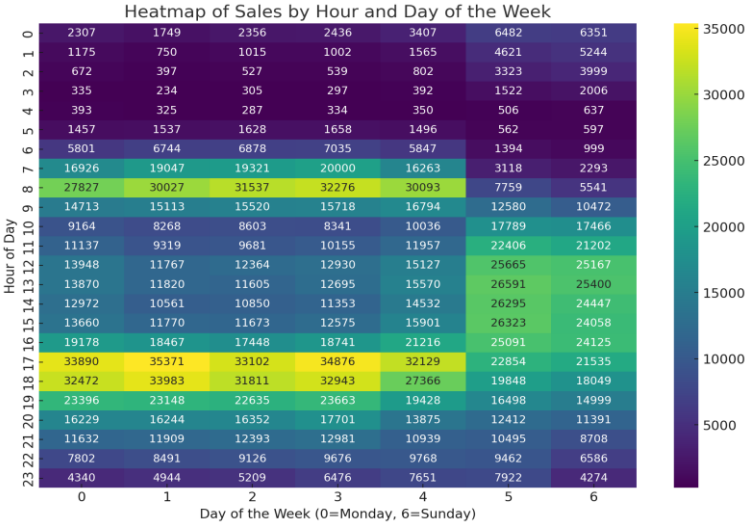


**Fig. 2.** Hourly Heatmap Analysis of Consumer Purchasing Behavior

The hourly heatmap analysis of consumer purchasing behavior, depicted in Figure 2, offers a detailed view of sales activities throughout the week. This heatmap shows variations in sales volume across different times and dates, with darker colors indicating higher sales. Notably, sales surge during midday and evening hours, aligning with typ-

ical shopping times after work or during lunch breaks. This pattern is especially pronounced on specific days like Friday, suggesting a peak in consumer activity as the weekend approaches.

### 4.2 Forecasting Performance Comparison

The performance comparison detailed in Table 1 reveals the CNN_BiLSTM_RF model's superiority over traditional and simpler deep learning architectures. This model demonstrates significantly lower MAE (14.4532) and RMSE (21.277), indicating a considerable reduction in prediction errors and stable performance under variable market conditions. Additionally, its exceptional $R^2$ value of 0.98751 and a superior MAPE of 0.35124 highlight its excellent fit and precision in capturing sales variability, making it a reliable tool for businesses needing accurate forecasts for inventory and supply chain management.

**Table 1.** Comparative Performance Metrics of the CNN_BiLSTM_RF Model Against Other Models

| Model | MAE | MAPE | RMSE | R2 |
|---|---|---|---|---|
| DT | 20.6304 | 0.4594 | 30.4869 | 0.97437 |
| LSBoost | 19.4895 | 0.35787 | 33.5362 | 0.96898 |
| CNN | 20.8123 | 0.51863 | 27.7471 | 0.97877 |
| LSTM | 30.9406 | 0.62332 | 46.9841 | 0.9591 |
| GRU | 20.9732 | 0.57491 | 30.1289 | 0.97496 |
| BiLSTM | 28.4718 | 0.55911 | 38.5243 | 0.95907 |
| CNN_LSTM | 28.2188 | 0.5421 | 38.9934 | 0.95807 |
| CNN_GRU | 25.817 | 0.37542 | 36.8379 | 0.96257 |
| CNN_BiLSTM | 24.588 | 0.60754 | 33.7661 | 0.96727 |
| CNN_BiLSTM_RF | 14.4532 | 0.35124 | 21.277 | 0.98751 |

Overall, the CNN_BiLSTM_RF model effectively integrates CNN's feature extraction capabilities, BiLSTM's temporal dependency insights, and Random Forests' generalization strengths. This combination addresses the complex challenges of time series forecasting, thereby enhancing prediction accuracy and stability in dynamic environments.

## 5 Conclusion

This study introduced a novel hybrid machine learning framework that effectively integrates CNN, BiLSTM, and RF to enhance sales forecasting accuracy and reliability. By synergistically combining the capabilities of CNN for high-resolution feature extraction and BiLSTM for capturing temporal dependencies with the robust generalization features of RF, the model addresses key challenges in sales data prediction, such as non-linear complexity and high volatility. Extensive experiments have demonstrated

that this approach significantly outperforms traditional and standalone machine learning models in predictive accuracy, establishing a new benchmark in the field of predictive analytics for sales. Future work will focus on optimizing the computational efficiency of the model and enhancing its robustness in handling datasets with sparse or irregular patterns to expand its applicability to a broader range of market conditions.

## Reference

1. Ma S, Fildes R. Retail sales forecasting with meta-learning[J]. European Journal of Operational Research, 2021, 288(1): 111-128.
2. Shetty S K, Buktar R. A comparative study of automobile sales forecasting with ARIMA, SARIMA and deep learning LSTM model[J]. International Journal of Advanced Operations Management, 2022, 14(4): 366-387.
3. Efat M I A, Hajek P, Abedin M Z, et al. Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales[J]. Annals of Operations Research, 2022: 1-32.
4. Jenkins G M, Box G E P. Time series analysis: forecasting and control[J]. (No Title), 1976.
5. Rob J. Hyndman, Athanasopoulos G. Forecasting: principles and practice[M]. Melbourne: OTexts, 2018.
6. Ahmed N K, Atiya A F, Gayar N E, et al. An empirical comparison of machine learning models for time series forecasting[J]. Econometric reviews, 2010, 29(5-6): 594-621.
7. Zhang G, Patuwo B E, Hu M Y. Forecasting with artificial neural networks: The state of the art[J]. International journal of forecasting, 1998, 14(1): 35-62.
8. Pai P F, Lin C S. A hybrid ARIMA and support vector machines model in stock price forecasting[J]. Omega, 2005, 33(6): 497-505.
9. Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. Journal of big Data, 2021, 8: 1-74.
10. Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction[J]. Neural Computing and Applications, 2021, 33(10): 4741-4753.
11. Zhang H, Zimmerman J, Nettleton D, et al. Random forest prediction intervals[J]. The American Statistician, 2020.