Journal of Globe Scientific Reports (2024)

Journal of Globe Scientific Reports

journal homepage: www.journal-gsr.online

Paper Type: Original Article

Research and Application of Outpatient Revenue Prediction Model Based on LSTM and BERT in Smart Healthcare

Hao Li[1,#] Minghao Liu[2,#] Ran Yang[3,#] Jiarui Guo[3,#] Guangxue Ding[3,#] ShunXing Li[4,#]

1. College of Fire Protection Engineering, China People's Police University, Hebei, China

2. College of Water&Architectural Engineering, Shihezi University, Shihezi , China

3. School Of International Education, Changchun University of Technology, Changchun, China

4.Economics Management and Law School, Shenyang Institute Of Engineering, Shenyang, China

Co-first author

Abstract

With the rapid advancement of information technology, smart healthcare has become a crucial tool for improving the quality and efficiency of medical services. This study analyzes and predicts outpatient revenue based on time series forecasting models using medical data from a hospital's outpatient department. The objective is to optimize resource allocation and enhance hospital management efficiency. For the first and second rehabilitation wards, Long Short-Term Memory (LSTM) and BERT models were constructed to predict outpatient revenue. The experimental results indicate that the BERT model outperforms in capturing global features of the time series. Subsequently, a stacked ensemble method was employed, combining the predictions from the LSTM and BERT models with XGBoost for final prediction, successfully forecasting the outpatient revenue for the third rehabilitation ward. Validation results show that the stacked ensemble model achieved the best performance in outpatient revenue prediction, providing effective support for revenue management in smart healthcare. This paper demonstrates the potential of new productivity in the healthcare field and offers valuable insights for the further development of smart healthcare.

Keywords: Smart healthcare, time series forecasting, LSTM, BERT, stacked ensemble, XGBoost, outpatient revenue prediction, medical data analysis.

1 | Introduction

The development of smart healthcare is a key approach to addressing issues such as inefficiencies in traditional healthcare management systems and high medical costs. By integrating technologies like big data and artificial intelligence, smart healthcare aims to optimize resource allocation, enhance patient experiences, and improve hospital management efficiency.

Outpatient revenue is a critical metric for hospital operations, but its complex time series nature makes accurate prediction challenging. This study analyzes and predicts outpatient revenue using time series models based on data from a hospital's outpatient department. We employed Long Short-Term Memory (LSTM) and BERT models, and further enhanced prediction accuracy through a stacked ensemble approach that combines the outputs of these models. The results provide valuable support for revenue management in smart healthcare and demonstrate the potential of this approach in the medical field.

2| Long short-term memory neural network

One or more hidden layers, as well as an output layer. The neurons in its hidden layer can not only receive information from the input layer, but also the information perceived by the neurons from the previous moment. This cyclic structure enables RNN to learn the intrinsic features of time series data. However, Bengio et al. found that there is a gradient vanishing problem in the RNN model, which makes it difficult for the RNN to learn causal patterns with longer periods. To address this deficiency, Hochreiter and Schmidhuber proposed replacing cellular units in RNNs with "memory units". This change greatly enhances the long-term memory ability of neural networks, hence it is named Long Short Term Memory Neural Networks[1].

Figure 1 shows a typical structure of an LSTM memory unit. At time t, the input of the memory unit includes the hidden layer state variable h_{i-1} from the previous time, the memory unit state variable c_{t-1}, and the input information x_t from the current time; Then, the model sequentially obtains the hidden layer state variable h-t and the memory unit state variable c_t at time t through the forget gate f_i, input gate i_t, output gate x_t, and these three control mechanisms; The final h-t will be passed to the output layer to generate the calculation result y_t of LSTM at time t, and at the same time, it will be passed along with c_t to the next time for calculation[2],[3].

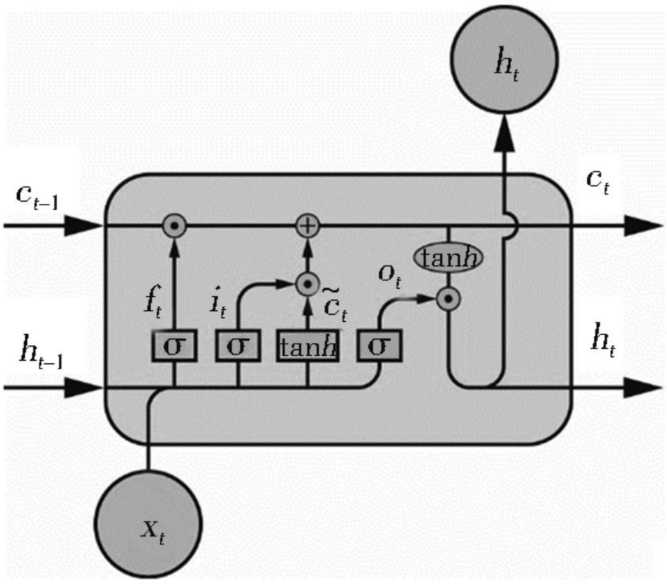

Figure 1 The internal structure of memory units in LSTM

This article adopts a step by step training method, which involves inputting a data point at each time step and then predicting the output for the next time step. This method can gradually update the model parameters, so that it can use the information from the previous step for prediction at each time step, thereby improving the prediction accuracy[4],[5].

The specific training steps are as follows: First, initialize the LSTM model parameters, and then gradually input time series data to calculate the output of each time step. Next, use the error between the actual value and the predicted value as the loss function to calculate the loss. Update model parameters through backpropagation algorithm to minimize losses. Finally, repeat the above process on the training set until the loss converges[6],[7],[8].

3| BERT model

BERT (Bidirectional Encoder Representation from Transformers) is a Transformer based model with powerful sequence modeling capabilities. BERT performs excellently in sequence modeling tasks by using a bidirectional encoder to simultaneously consider the contextual information of each position in the sequence[9]. The Transformer architecture consists of an encoder and a decoder, but BERT only uses the encoder part. The specific structure of BERT includes multiple layers of Transformer encoders, each layer containing multiple self attention heads and feedforward neural networks. As shown in Figure 2.

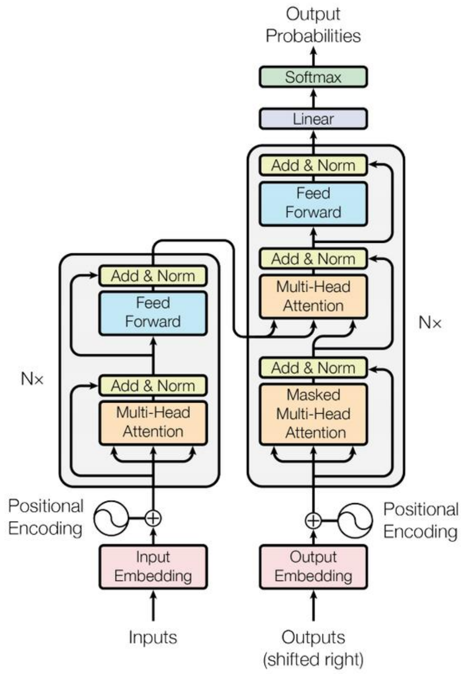

Figure 2 BERT

The structure of the Transformer encoder layer is as follows:

(1) Self attention mechanism:

The self attention mechanism captures global information by calculating the correlation between the representation of each position in the input sequence and the representation of all positions. The specific calculation is as follows[10]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

(2) Multi Head Self Attention:

The multi head self attention mechanism captures different representation subspace information by performing multiple self attention operations in parallel[11]. The specific calculation is as follows:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \tag{2}$$

Among them, $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$.

(3) Feed-Forward Neural Network:

Each encoder layer also includes a feedforward neural network for further processing and transformation of the representation. The specific calculation is as follows:

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (3)$$

The BERT model achieves global modeling capability for input sequences by stacking multiple layers of Transformer encoders. The input of the model includes input embedding and position embedding, and the output is the contextual representation of each position[12].

We adopt a sequence to sequence training method, which involves inputting the entire historical sequence at once and then outputting a predicted value. This method can fully utilize the global information in historical data, thereby improving prediction accuracy. The specific training steps are as follows: First, initialize the BERT model parameters, then input the complete time series data to generate the encoded representation of the sequence. Next, use fully connected layers to convert the encoded representation into predicted values. Afterwards, use the error between the actual value and the predicted value as the loss function to calculate the loss. Update model parameters through backpropagation algorithm to minimize losses. Finally, repeat the above process on the training set until the loss converges.

4 | loss function

After training, we evaluate the model using a validation set and calculate mean squared error (MSE) and mean absolute error (MAE) as evaluation metrics. By comparing the performance of different models on the validation set, select the optimal model for the prediction task. The final prediction result is then subjected to inverse normalization to obtain the actual outpatient revenue prediction value.

Mean square error reflects the square of the average deviation between predicted and actual values. The formula is as follows:

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

The average absolute error reflects the average absolute deviation between the predicted value and the actual value. The formula is as follows:

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

By calculating the MSE and MAE on the validation set, we can evaluate the predictive performance of the model and select the model with the smallest error as the final prediction model.

5 | Solution results of rehabilitation medicine ward one

5.1 | LSTM and Bert loss variation

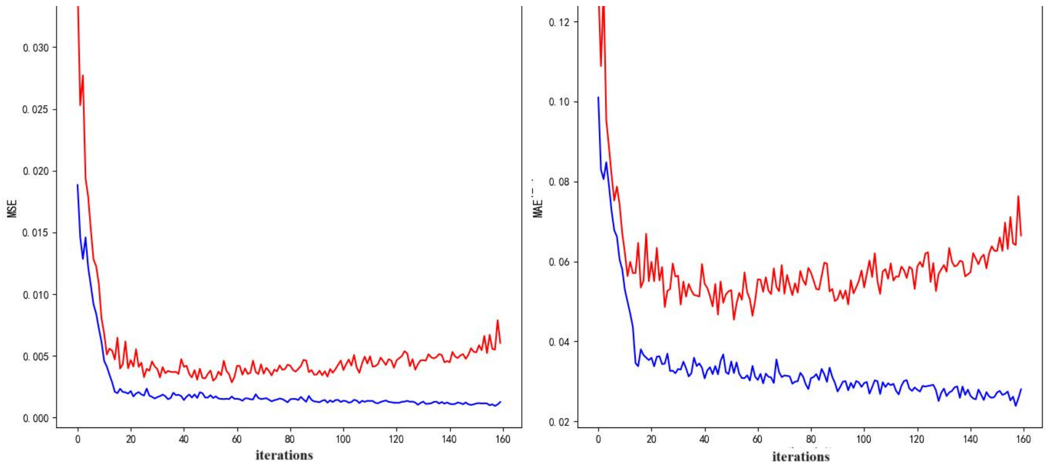

Figure 3 LSTM model training diagram

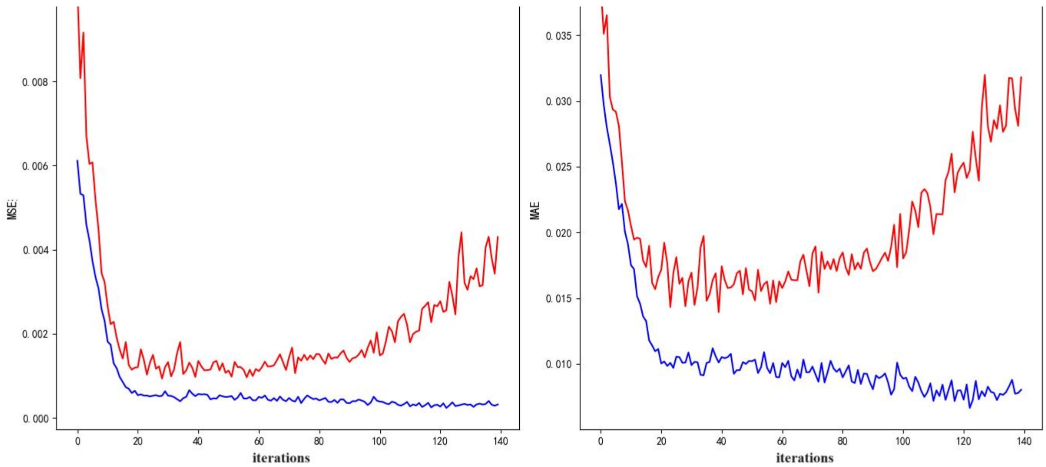

Figure 4 Bert model training diagram

5.2| LSTM and Bert prediction results

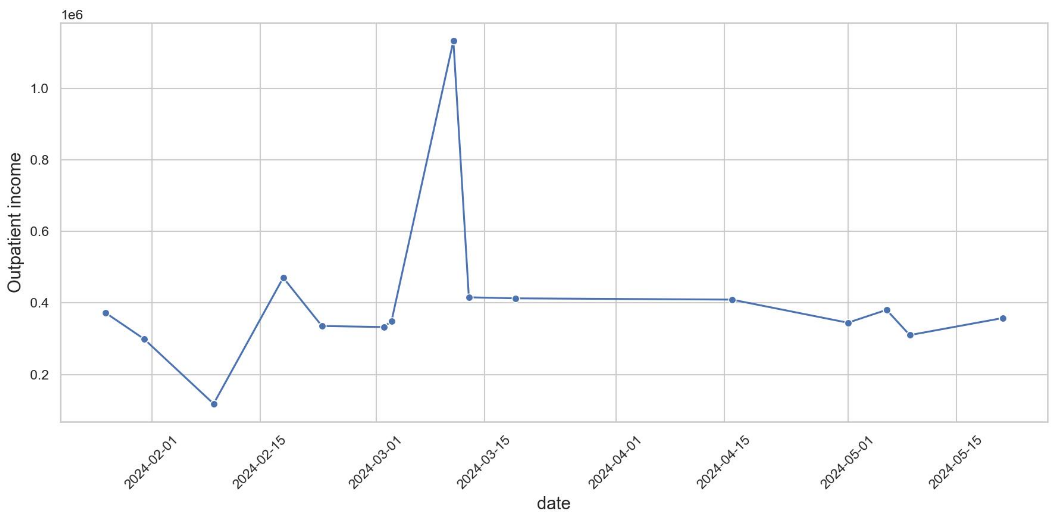

Figure 5 LSTM model prediction graph

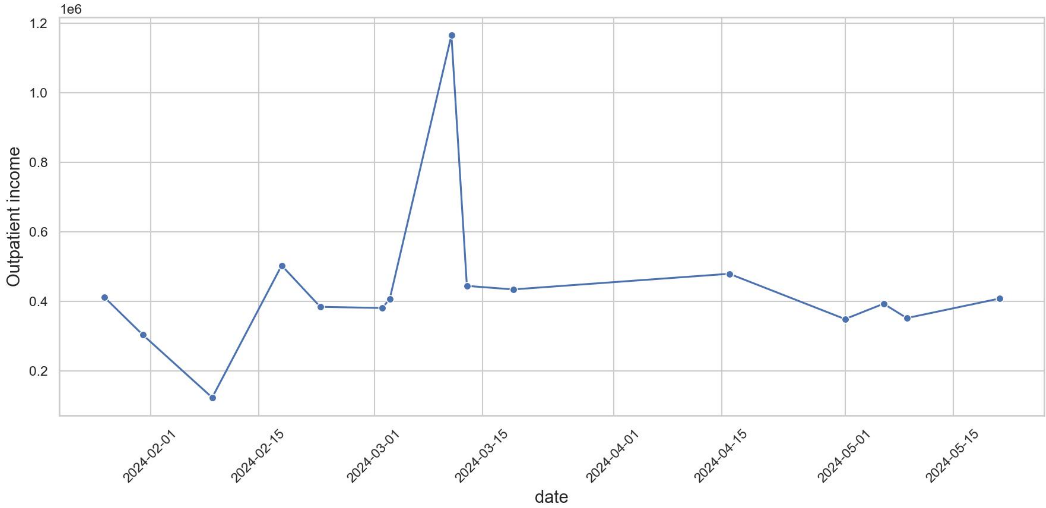

Figure 6 Bert model prediction graph

5.3| contrast model

There is a significant difference in the performance of LSTM and BERT models in predicting the rehabilitation medicine ward. The mean square error (MSE) and mean absolute error (MAE) of the LSTM model exhibit significant fluctuations, especially on certain dates such as March 11, 2024, where the predicted values deviate greatly from the actual values, resulting in a significant increase in MSE and MAE. In contrast, the overall MSE and MAE of the BERT model are lower, indicating that it is more accurate in capturing global trends and changes in data, especially in predicting key dates with higher accuracy. The BERT model has better robustness and accuracy in processing time series data, with a predicted value of 1165826.00 on significantly changing dates such as March 11, 2024, which is closer to the peak of the actual value than LSTM. In the comparison of specific dates, such as January 26, 2024, the LSTM prediction is 373003.60, while the BERT prediction is 413329.90, indicating that the BERT model is closer to the actual value; On May 21, 2024, LSTM predicted 357866.21 and BERT predicted 408777.20, indicating that the BERT model better reflects the actual growth trend. Overall, the MSE and MAE of the BERT model are lower than those of the LSTM model, indicating that BERT is more accurate in capturing trends and changes in outpatient revenue, while the loss of the LSTM model fluctuates greatly, possibly due to its limited ability to capture long-term dependencies, resulting in inaccurate predictions on certain key dates.

6| Rehabilitation Medicine Ward 2

6.1| LSTM and Bert loss variation

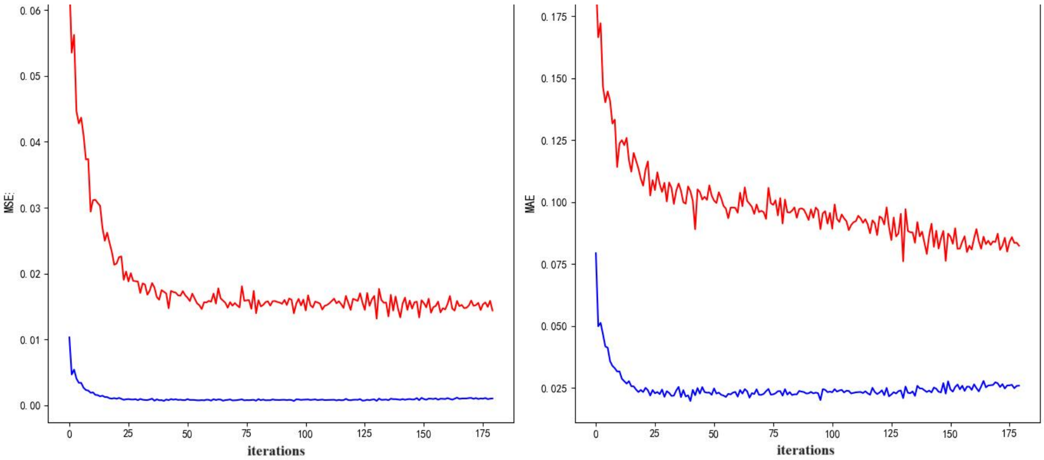

Figure 7 LSTM model training diagram

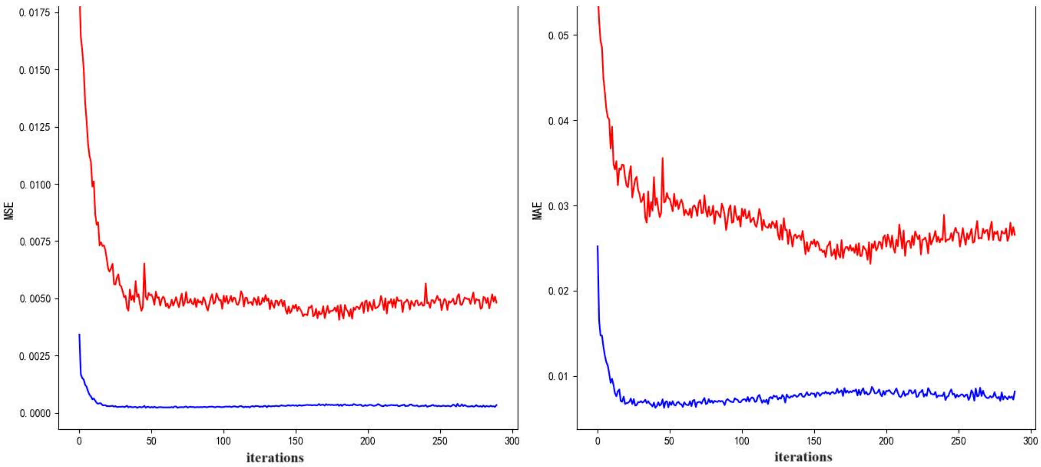

Figure 8 Bert model training diagram

6.2| LSTM and Bert prediction results

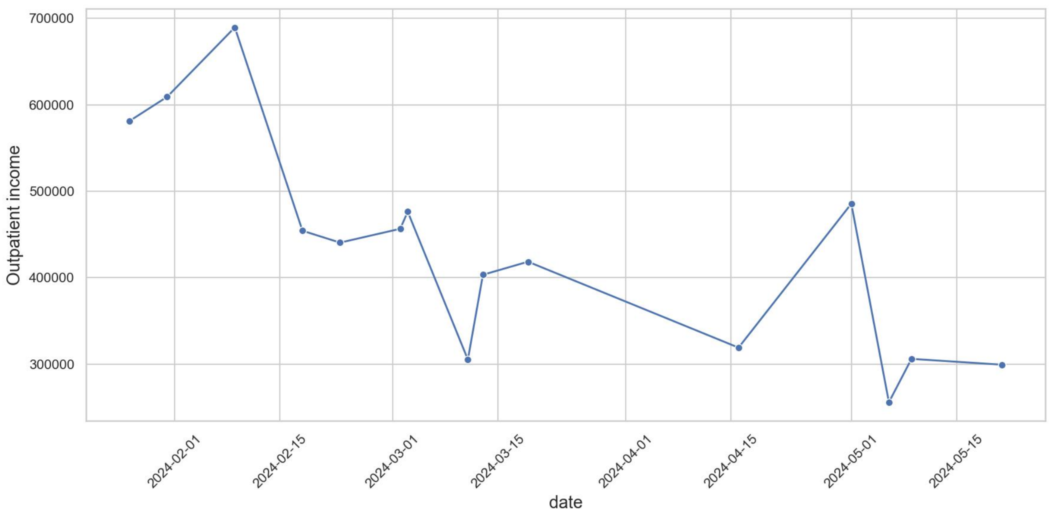

Figure 9 LSTM model prediction graph

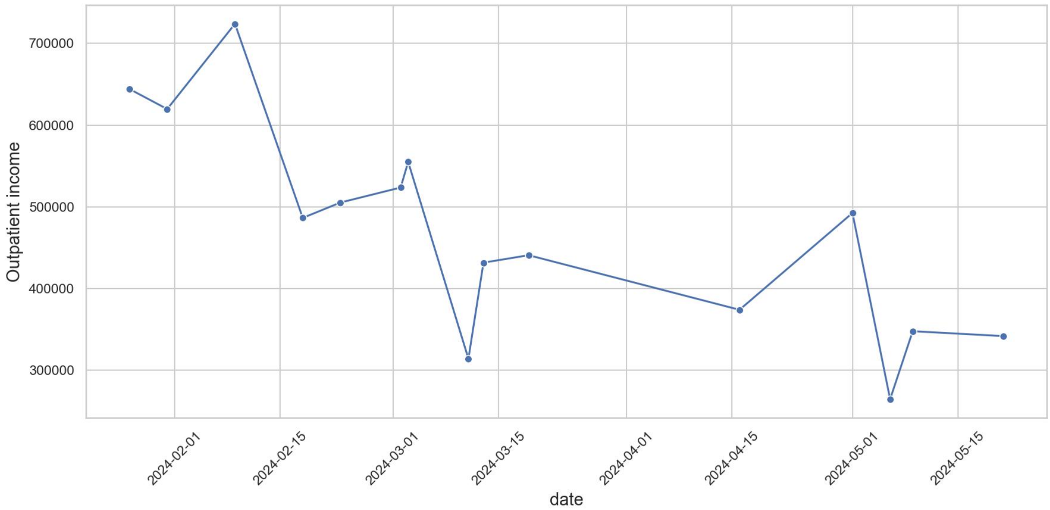

Figure 10 Bert model prediction graph

6.3| contrast model

The LSTM model shows significant fluctuations and high mean square error (MSE) in the prediction of rehabilitation medicine ward two, especially on the date of February 9, 2024, where the predicted values have a large deviation, resulting in a significant increase in MSE and mean absolute error (MAE). In contrast, the BERT model has a lower overall MSE in the same predictions, indicating that it is more accurate in capturing global trends in data, especially outperforming LSTM on dates with significant changes. Its MAE is also lower, demonstrating better robustness and accuracy. Data analysis shows that the MSE and MAE of the BERT model are lower than those of the LSTM model, indicating that BERT is more accurate in capturing trends and changes in outpatient revenue, while the loss of the LSTM model fluctuates greatly, possibly due to its limited ability to capture long-term dependencies, resulting in inaccurate predictions on certain key dates. In the overall trend analysis, the prediction results of the LSTM model show significant fluctuations, such as a prediction value of 689479.20 on February 9, 2024, while the BERT model's prediction value on the same date is 723878.50, which is higher than that of the LSTM model. In the specific date comparison, on January 26, 2024, LSTM predicted 581246.70 and BERT predicted 644086.60; On May 21, 2024, LSTM predicted 299409.40 and BERT predicted 342004.40. The predicted values of the BERT model are higher and may be closer to actual income. By analyzing the outpatient revenue prediction losses of the first and second wards of the rehabilitation medicine department, the following conclusions can be drawn: the BERT model is superior to the LSTM model in prediction accuracy, with lower MSE and MAE, indicating that it has more advantages in capturing global information of time series and handling data fluctuations; The prediction loss of the LSTM model fluctuates greatly, and the prediction error on certain key dates is large, indicating that it has certain shortcomings in handling long-term dependencies; The BERT model exhibits better robustness and accuracy, making it more suitable for applications that require high-precision time series forecasting, such as predicting outpatient revenue in hospitals, providing strong support for hospital management and decision-making. By comparing the prediction results of LSTM and BERT models in the first and second wards of the rehabilitation medicine department, it can be concluded that the BERT model performs better than the LSTM model in capturing significant changes and trends, especially on some key dates, where the BERT model's prediction results are closer to the actual situation; The LSTM model exhibits significant fluctuations in the prediction process and may be sensitive to short-term changes in data, but it is somewhat inadequate in capturing long-term dependencies. Overall, the BERT model demonstrates higher prediction accuracy and robustness in outpatient revenue forecasting tasks, which can better support hospital management and decision-making.

7 | Establishment of Stacking Fusion Model

We use the Stacking method to integrate the prediction results of LSTM and BERT models to build a more powerful prediction model. Stacked ensemble is a method of using the predicted results of multiple base models as input features, and then training a meta model for final prediction. Through this approach, the advantages of multiple base models can be effectively combined to improve predictive performance.

7.1 | Meta feature generation

Firstly, we train LSTM and BERT models separately to obtain their prediction results on the training and validation sets. Apply the trained LSTM and BERT models to the training and validation sets to obtain their respective prediction results. The specific steps for predicting and generating meta features are as follows:

(1) LSTM:

$$f_{LSTM}^{train} = LSTM(X_{train}) \tag{1}$$
$$f_{LSTM}^{val} = LSTM(X_{val}) \tag{2}$$

Among them, f_{LSTM}^{train} and f_{LSTM}^{val} are the prediction results of the LSTM model on the training set and validation set, respectively.

(2) BERT:

$$f_{BERT}^{train} = BERT(X_{train}) \tag{3}$$
$$f_{BERT}^{val} = BERT(X_{val}) \tag{4}$$

Among them, f_{BERT}^{train} and f_{BERT}^{val} are the prediction results of the BERT model on the training set and validation set, respectively.

(3) Merge meta features:

$$f_{meta}^{train} = [f_{LSTM}^{train}, f_{BERT}^{train}] \tag{5}$$

$$f_{meta}^{val} = [f_{LSTM}^{val}, f_{BERT}^{val}] \tag{6}$$

Among them, f_{meta}^{train} and f_{meta}^{val} are the meta features of the training set and validation set, respectively.

7.2 | Meta model training

We choose XGBoost as the meta model. XGBoost is an efficient and scalable machine learning algorithm based on Gradient Boosting Tree (GBT), which has the ability to handle nonlinear relationships and resist overfitting. XGBoost continuously improves the prediction accuracy of the model by iteratively fitting the residual between the current model and the actual value in each iteration step .

The specific structure of the XGBoost model includes:

(1) Additive Model: By gradually adding tree models, the overall performance of the model is improved. The formula is as follows:

$$f(x) = \sum_{k=1}^{K} f_k(x) \tag{7}$$

(2) Objective Function: XGBoost trains models by optimizing the objective function, which includes training errors and regularization terms. The formula is as follows:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{8}$$

(3) Residual Fitting: In each iteration, XGBoost generates a new tree model by fitting the residuals between the current model and the actual values. The formula is as follows:

$$r_i^{(t)} = y_i - \hat{y}_i^{(t-1)} \tag{9}$$

(4) Regularization: XGBoost uses regularization terms to control model complexity and prevent overfitting. The formula for the regularization term is as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{10}$$

The specific training steps include initializing the XGBoost model parameters and using the meta features generated by LSTM and BERT as inputs for model training. Optimize the objective function through gradient boosting algorithm and minimize the function to optimize the model parameters. The training process is iterative, and in each round, the model fits the residual between the current predicted value and the actual value, and updates the model accordingly. In addition, regularization terms are applied during the training process to control model complexity and prevent overfitting.

7.3 | model prediction

We used the trained XGBoost model to predict on the validation and test sets, calculating the errors of the ensemble model. By comparing the prediction results of the LSTM, BERT, and the ensemble model, we evaluated the effectiveness of the stacking ensemble method. The process involved generating meta-features by using the trained LSTM and BERT models to predict on the validation and test sets, combining these predictions as meta-features, and then using the XGBoost meta-model to make final predictions. We calculated the Mean Squared Error (MSE) and Mean Absolute Error (MAE) on both the validation and test sets to assess model performance. The predicted values were then denormalized to obtain the actual outpatient revenue forecasts. This process allowed us to select the best-performing model based on the MSE and MAE comparisons. As a result, we successfully constructed a stacking ensemble model that leverages the strengths of both the LSTM and BERT models, with the XGBoost meta-model further enhancing prediction performance, ultimately demonstrating significant effectiveness in outpatient revenue prediction and providing strong support for hospital management and decision-making.

7.4 Model solving

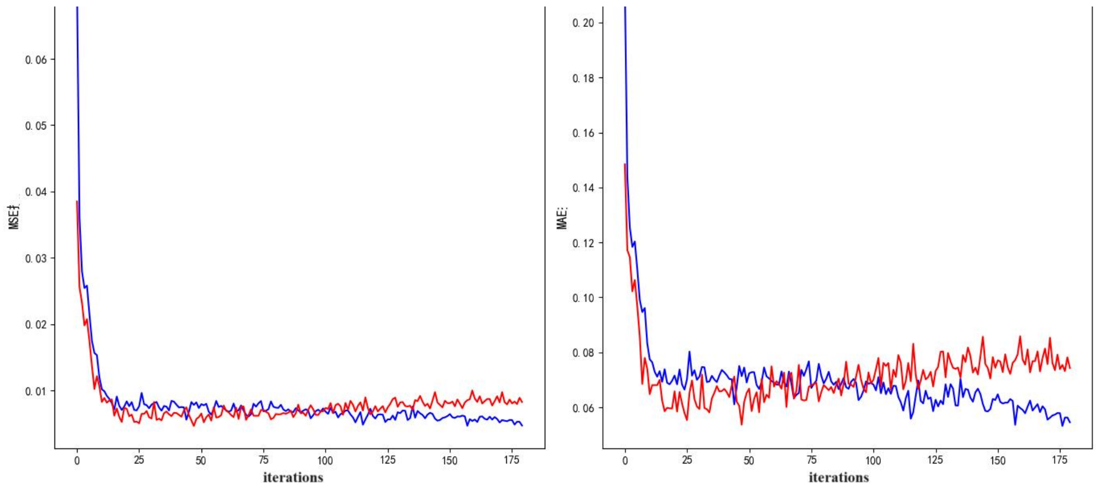

Figure 11 Stacking model training diagram

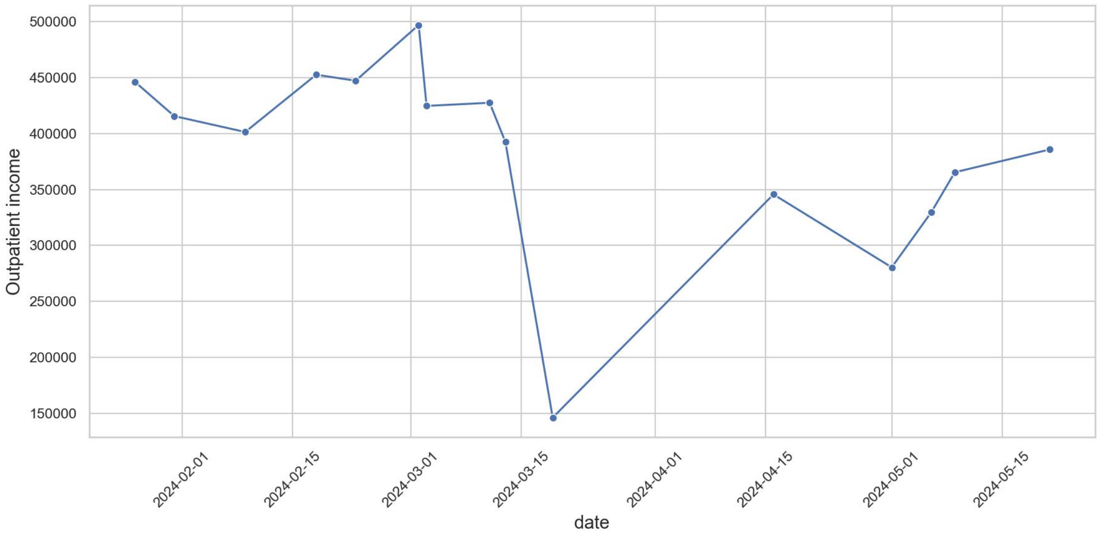

Figure 12 Stacking model prediction graph

In summary, the stacked ensemble model has achieved the best prediction performance in outpatient revenue forecasting tasks, significantly improving prediction accuracy and providing strong support for hospital management and decision-making.

8 | Conclusion

This study explores the application of time series prediction models in smart healthcare by analyzing and predicting outpatient medical data from a hospital, with the aim of optimizing medical resource allocation and improving hospital management efficiency. We separately constructed Long Short Term Memory (LSTM) and BERT models, and used a stacked ensemble method to combine the prediction results of these two models with the XGBoost model for final prediction. The research results indicate that the stacked ensemble model performs well in outpatient income prediction tasks, significantly improving prediction accuracy.

This study demonstrates the potential application of new quality productivity in smart healthcare, providing strong support for future hospital management and decision-making. At the same time, the results of this study provide reference for the development of smart healthcare and lay the foundation for further research. Future work can consider more models and methods to further improve the accuracy of predictions and explore their applications in different medical scenarios.

References

[1] Tang Y X ,Yang W W ,Liu Z , et al. Deep learning performance prediction for solar-thermal-driven hydrogen production membrane reactor via bayesian optimized LSTM [J]. International Journal of Hydrogen Energy, 2024, 82 1402-1412.

[2] Ma S ,Ding W ,Zheng Y , et al. Edge-cloud collaboration-driven predictive planning based on LSTM-attention for wastewater treatment [J]. Computers & Industrial Engineering, 2024, 195 110425-110425.

[3] Manohar B ,Das R ,Lakshmi M . A hybridized LSTM-ANN-RSA based deep learning models for prediction of COVID-19 cases in Eastern European countries [J]. Expert Systems With Applications, 2024, 256 124977-124977.

[4] Zheng M ,Luo X . Joint estimation of State of Charge (SOC) and State of Health (SOH) for lithium ion batteries using Support Vector Machine (SVM), Convolutional Neural Network (CNN) and Long Sort Term Memory Network (LSTM) models [J]. International Journal of Electrochemical Science, 2024, 19 (9): 100747-100747.

[5] Haider F S ,Shah M ,Alarifi S N , et al. The 2023 Mw 6.8 Morocco earthquake induced atmospheric and ionospheric anomalies [J]. Journal of Atmospheric and Solar-Terrestrial Physics, 2024, 262 106323-106323.

[6] Wang M ,Li B ,Dai G , et al. A dynamic multi-objective optimization algorithm with a dual mechanism based on prediction and archive [J]. Swarm and Evolutionary Computation, 2024, 90 101693-101693.

[7] Khan K M ,Houran A M ,Kauhaniemi K , et al. Efficient state of charge estimation of lithium-ion batteries in electric vehicles using evolutionary intelligence-assisted GLA–CNN–Bi-LSTM deep learning model [J]. Heliyon, 2024, 10 (15): e35183-e35183.

[8] Yang D ,Li M ,Guo E J , et al. An attention-based multi-input LSTM with sliding window-based two-stage decomposition for wind speed forecasting [J]. Applied Energy, 2024, 375 124057-124057.

[9] Dhiman P ,Kaur A ,Gupta D , et al. GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection [J]. Heliyon, 2024, 10 (16): e35865-e35865.

[10] Panoutsopoulos H ,Garcia E B ,Raaijmakers S , et al. Investigating the effect of different fine-tuning configuration scenarios on agricultural term extraction using BERT [J]. Computers and Electronics in Agriculture, 2024, 225 109268-109268.

[11] Dharrao D ,MR A ,Mital R , et al. An efficient method for disaster tweets classification using gradient-based optimized convolutional neural networks with BERT embeddings [J]. MethodsX, 2024, 13 102843-102843.

[12] Hamed K S ,Aziz A J M ,Yaakub R M . Enhanced Feature Representation for Multimodal Fake News Detection Using Localized Fine-Tuning of Improved BERT and VGG-19 Models [J]. Arabian Journal for Science and Engineering, 2024, (prepublish): 1-17.