

A transformer-based framework for enterprise sales forecasting

Yupeng Sun¹ and Tian Li²

¹ School of Accounting, Yunnan University of Finance and Economics, Yunnan, China

² School of Accounting, Tianjin University of Commerce, Tianjin, China

ABSTRACT

Sales forecasting plays an important role in business operations as it impacts decisions on inventory management, allocation of resources, and financial planning. Accurate sales predictions are essential for optimizing cash flow management, adapting marketing and sales strategies, and facilitating strategic planning. This study presents a computational framework for predicting business sales using transformers, which are considered one of the most powerful deep learning architectures. The design of our model is specifically tailored to accommodate tabular data with low dimensions. The experimental results demonstrated that our proposed method surpasses conventional machine learning models, achieving reduced mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE), as well as higher R^2 values of nearly 0.95. The results confirmed that the model is applicable not only to this research but also to similar studies that use low-dimensional tabular data. The improved accuracy and stability of our model demonstrate its potential as a useful tool for enhancing sales prediction, therefore facilitating more informed decision-making and strategic planning in corporate operations.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Data Science, Neural Networks

Keywords Sales forecasting, Transformers, Deep learning, Business intelligence

INTRODUCTION

Background

Sales forecasting plays a vital role in business activities from profit estimation to decision-making (Lawrence, O'Connor & Edmundson, 2000; Wacker & Lummus, 2002). Effective sales forecasting shapes decisions related to inventory control (Fildes & Beard, 1992; Snyder, Koehler & Ord, 2002; Snyder, 2002), resource allocation (Wacker & Lummus, 2002; Gupta & Kohli, 2006; Sugiarto, Sarno & Sunaryono, 2016), and financial planning (Osadchiy, Gaur & Seshadri, 2013; Zhu, 2023; Xu, Li & Donta, 2024). Sales history data, in combination with socioeconomic factors such as current market dynamics, economic indicators, consumer trends, and seasonal impacts, are typically used to predict sales for upcoming periods (Gao et al., 2014; Cheriyan et al., 2018). Effective forecasts are crucial for optimizing cash flow management (Dharan, 1987) and adjusting marketing (Biswas, Sanyal & Mukherjee, 2023) and sales strategies (Janczura & Michalak, 2020; Zhu, Bahadur & Ali, 2023). An increase in demand might lead to ramped-up production or enhanced promotional activities to maximize sales, while accurate forecasting of a sales decline might

Submitted 14 August 2024
Accepted 21 October 2024
Published 21 November 2024

Corresponding author
Yupeng Sun, syupeng@126.com

Academic editor
Hoang Nguyen

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj-cs.2503

© Copyright
2024 Sun and Li

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

prompt cost-cutting measures (Mccarthy et al., 2006). Furthermore, sales forecasting also serves as a cornerstone of strategic planning, helping to set achievable sales goals that synchronize the efforts of marketing, sales, production, and logistics teams (Davis & Mentzer, 2007). Improved forecasting accuracy enhances customer satisfaction by ensuring that businesses can meet demand promptly and reliably (Lyu & Choi, 2020; He, 2022; Ban et al., 2023). Therefore, improving sales forecasting techniques is one of the major approaches to maintain a business's operational efficiency and financial stability, allowing it to adapt quickly to changing market conditions and consumer preferences, thereby sustaining growth and competitiveness (Lawrence & O'Connor, 2000).

In the era of artificial intelligence (AI) (Sohrabpour et al., 2021) and big data (Thomassey & Zeng, 2018), sales forecasting has undergone a transformative shift, harnessing advanced technologies to predict future sales with unprecedented precision. AI algorithms and machine learning (ML) models are now central to analyzing vast amounts of data, identifying patterns and trends that were previously indiscernible (Chen & Lu, 2016). This has allowed businesses to refine their forecasts beyond traditional statistical modeling methods, integrating real-time data from a variety of sources including market shifts, consumer behavior online, social media sentiments, and even weather patterns (Ramos, Santos & Rebelo, 2015; Leow, Nguyen & Chua, 2021). The integration of big data analytics into sales forecasting provides a holistic view of the marketplace (Boone et al., 2019). Companies can now process and analyze data from diverse and extensive datasets instantaneously. This ability enables more dynamic and agile responses to market changes, as AI can quickly adapt forecasts in response to promotional campaigns or sudden shifts in consumer demand, providing businesses with the agility to optimize inventory levels and tailor marketing strategies effectively (Forrest & Hoanca, 2015). Moreover, AI-driven tools offer predictive insights with a level of granularity that allows companies to segment customers more precisely, tailor products to specific markets, and adjust prices dynamically (Chen et al., 2024). The predictive power of AI also extends to identifying potential market opportunities and risks before they fully emerge, giving companies strategic advantages. Overall, the advent of AI and big data has not only enhanced the accuracy of sales forecasts but has also revolutionized how businesses strategize and operate (Thomassey & Zeng, 2018; Sohrabpour et al., 2021). As these technologies continue to evolve, they promise to unlock even deeper insights and drive smarter, data-driven decision-making across industries. This advancement is important for businesses aiming to remain competitive in a rapidly changing economic landscape (Weber & Schütte, 2019).

Related works

For decades, numerous computational frameworks employing machine learning and deep learning have been developed for sales forecasting (or prediction). Chu & Zhang (2003) conducted a comparative study to compare the performance of linear and nonlinear models for retail sales forecasting. Das & Chaudhury (2006) developed a model to estimate the sales fluctuations of a footwear company over a period of time using recurrent neural networks. Their model was designed to predict weekly retail sales to minimize the uncertainty in short-term sales planning. Ni & Fan (2011) proposed a forecasting model

for the fashion retail using real-time data, combining two types of prediction: long-term and short-term. [Beheshti-Kashi et al. \(2014\)](#) implemented various state-of-the-art methods to construct forecasting models for fashion and new products under different modeling strategies. [Kaneko & Yada \(2016\)](#) utilized simple deep neural networks to develop a prediction model for retail store sales. [Ribeiro, Seruca & Durão \(2017\)](#) created a model specifically for predicting sales in a pharmaceutical distribution firm using exponential smoothing time-series. Their research tackled two major issues: exploring inventory allocation strategies to prevent stock shortages and forecasting sales to maintain sufficient levels of medicine inventory. [Punam, Pamula & Jain \(2018\)](#) introduced a two-level statistical model for forecasting big mart sales using linear regression, support vector regression, and cubist. [Tsoumakas \(2018\)](#) investigated a series of machine learning techniques used for food sales prediction. [Kohli, Godwin & Urolagin \(2020\)](#) used linear regression and k -nearest neighbors regression to develop sales prediction models. [Yao \(2023\)](#) employed three machine learning algorithms, including decision tree, random forest, and k -neighbors regression, to build a forecasting model for Walmart sales prediction.

Motivations

Since most deep learning architectures used for sales forecasting are relatively simple, the potential of deep learning, especially more advanced models, has not been fully exploited. In our study, we developed a computational framework for enterprise sales forecasting using a streamlined transformer for low-dimensional tabular data. Our model draws inspiration from the TabTransformer ([Huang et al., 2020](#)), with minor modifications. As a transformer-based architecture specifically designed for tabular data, it performs more effectively on this type of data compared to traditional neural networks. Furthermore, all transformer-based architectures are characterized by attention mechanisms, which have been recognized as exceptionally powerful in recent years ([Vaswani et al., 2017](#); [Huang et al., 2020](#); [Badaro, Saeed & Papotti, 2023](#); [Mao, 2024](#)). Transformers and their variants have been widely used in natural language processing. After training, our developed models are rigorously assessed and compared against multiple conventional machine learning and deep learning algorithms. Additionally, model stability is evaluated by repeating experiments across multiple random trials.

PROPOSED MODEL

Model architecture

[Figure 1](#) visualizes the model architecture proposed in our study. The model receives two types of input features: categorical and numerical features. The categorical features are first passed through a Column Embedding layer (see subsection *Column Embedding*) before entering a Multi-head Attention layer. The embedding vectors are loaded into a Multi-head Attention layer, summed with the embedding residual, and normalized. The summed attention outputs continue to enter a Feed Forward layer. The Feed Forward layer's outputs are then summed with the attention residual and normalized to create categorical attention outputs. The numerical features are normalized before being processed by the

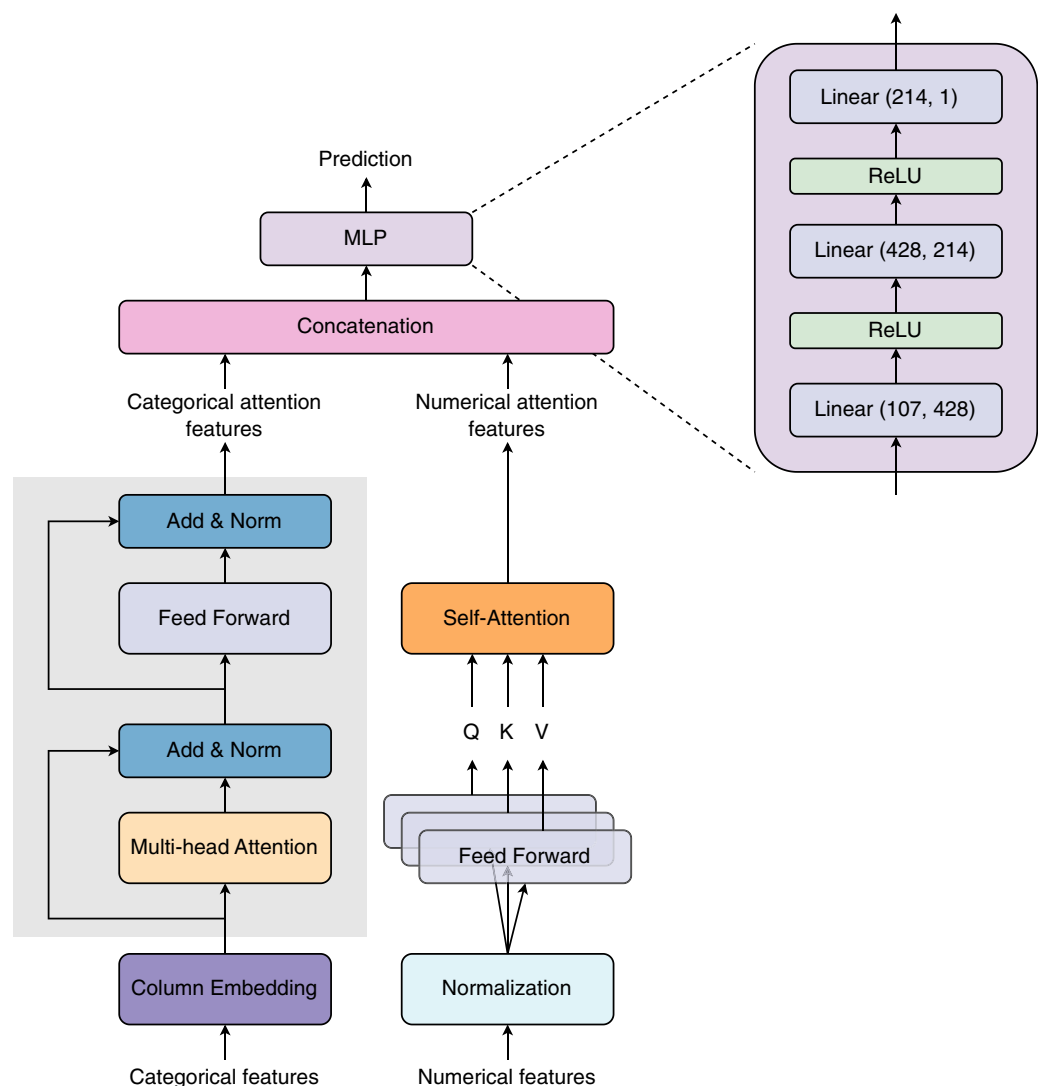


Figure 1 Model architecture proposed in our study. The network is designed with two major input blocks: Column Embedding and Normalization, for processing categorical (discrete or string) and numerical (continuous) features, respectively. Both types of features are passed through distinct attention layers for selective learning of important features. The selected attention features are then concatenated before being transferred to the Multilayer Perceptron block.

Full-size DOI: 10.7717/peerj-cs.2503/fig-1

three Feed Forward layers. The Feed Forward layers' outputs are used as Query (Q), Key (K), and Value (V) vectors for the Self-Attention layer. The Self-Attention layer learns numerical features *via* Q, K, and V vectors to create numerical attention outputs. The categorical and numerical attention outputs are concatenated and then transferred to a Multi-layer Perceptron (MLP) block to predict final outcomes. Multi-layer Perceptron block is specified by three linear layers. The input size of these linear layers is 107, 428, and 214, respectively. The output size of the prior linear layers is equal to the input size of the posterior layers. The Rectified Linear Unit (ReLU) is used as the activation function. In the TabTransformer architecture, [Huang et al. \(2020\)](#) applied one layer of normalization for

the numerical features branch. In contrast, we added an attention layer after the normalization layer for these features.

Column embedding

Column embedding is a technique specifically designed for modeling tabular data, utilizing the strengths of transformer models to address the unique challenges of categorical and numerical features in these datasets (Huang et al., 2020). In this approach, categorical features (columns) are transformed into dense vector representations, or embeddings, which allow the model to effectively learn from them. These embeddings capture both the identity and the contextual relationships between categories, which is crucial for understanding feature interactions that heavily influence the target variable (Mao, 2024). Numerical features are also incorporated alongside categorical embeddings, enabling the model to leverage both types of data in its predictions. The embedding process is typically achieved through learned embeddings, where each category is mapped to a vector in a continuous space. This allows the model to identify hidden relationships between categories that may not be obvious in the raw data. Furthermore, the contextual embeddings learned by the model are resilient to missing and noisy data, ensuring robust performance in real-world scenarios where data quality can vary. By integrating column embeddings, the model consistently outperforms traditional deep learning methods for tabular data, delivering higher predictive accuracy and improved generalization across a wide range of datasets (Huang et al., 2020; Mao, 2024).

Self-attention

The self-attention mechanism, derived from the transformer architecture, calculates attention scores to determine the relevance of each feature in relation to the others in the input data. This enables the model to concentrate on critical feature interactions that are essential for accurate predictions. Given an input matrix of size $X \in \mathbb{R}^{n \times d}$, n represents the number of features and d is the dimension of the embedding. The attention mechanism can be computed based on three fundamental components: Query (Q), Key (K), and Value (V). The Q, K, and V matrices are calculated through learnable weight matrices W_Q , W_K , and $W_V \in \mathbb{R}^{d \times d_k}$, respectively where d_k is the key dimension. The attention scores are computed by taking the dot product between the Q and K matrices, scaled by the square root of the key dimension d_k , as follow:

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

The dot product QK^T calculates the relevance between each query and key, while scaling by $\sqrt{d_k}$ helps maintain stable gradients when d_k is large. The softmax function then converts the attention scores into probabilities for better interpretation.

DATASET DESCRIPTION

In our study, we utilized the dataset from the Kaggle competition “Walmart Recruiting-Store Sales Forecasting” (Kaggle, 2014). This dataset consists of three dataframes, which we

Table 1 Information on datasets used for model training, validation, and testing.

Dataset	Time	Number of samples	Number of weeks
Training	05/02/2010 to 30/03/2012	331,742	113
Validation	06/04/2012 to 13/07/2012	44,192	15
Test	20/07/2012 to 26/10/2012	44,278	15
Total	05/02/2010 to 26/10/2012	420,212	143

merged to form a single dataframe containing 14 variables. These include 11 continuous variables: *Size*, *Temperature*, *Fuel_Price*, *CPI*, *Unemployment*, *Total_MarkDown*, *max*, *min*, *mean*, *median*, and *std*; along with three categorical variables: *Store*, *Dept*, and *Type*. The *Weekly_Sales* variable is labels for prediction. Since the dataset contains time-series samples, we created three sets of data: training, validation, and test sets based on the time order. The training set contains 331,742 samples from 05/02/2010 to 30/03/2012, equivalent to 113 weeks. The validation and test sets have 44,192 and 44,278 samples, respectively, both equivalent to 15 weeks. The validation data were recorded from 06/04/2012 to 13/07/2012, while the test data were recorded from 20/07/2012 to 26/10/2012. The whole dataset contains data recorded from 05/02/2010 to 26/10/2012, equivalent to 143 weeks with 420,212 samples. Table 1 summarizes information on datasets used for model training, validation, and testing. The data preprocessing was executed using module `preprocessing`, `scikit-learn` library (Pedregosa et al., 2011) version 1.4.1 in Python 3.11.8 environment.

EXPERIMENTS

Overview of modeling strategy

Unlike normal datasets whose samples can be randomly split, the time-series dataset contains ordered samples which cannot be reversed. Therefore, the training process was designed to direct all models to learn data sequentially. The initial training set contained samples of 113 weeks, and the validation set contained samples of 15 weeks, which were used to develop all models, including ours. Figure 2 visualizes the modeling strategy in our study. This strategy includes two phases: (i) *training and validation* and (ii) *training and testing*.

In iteration 1, the models were trained with data corresponding to 113 weeks and validated by data from week 1 (in the next 15 weeks). In iteration 2, the week-1 validation data were accumulated in the initial training set for updating the models by re-training them with the updated training set. This process was iteratively repeated until the models were validated by the data from week 15. During the *training and validation* phase, the models were retrained 15 times.

After completing the *training and validation* phase, we conducted the *training and evaluation* phase. In the second phase, the models were also created, updated, and evaluated in the same manner as the first phase. Since the test set contains data from the

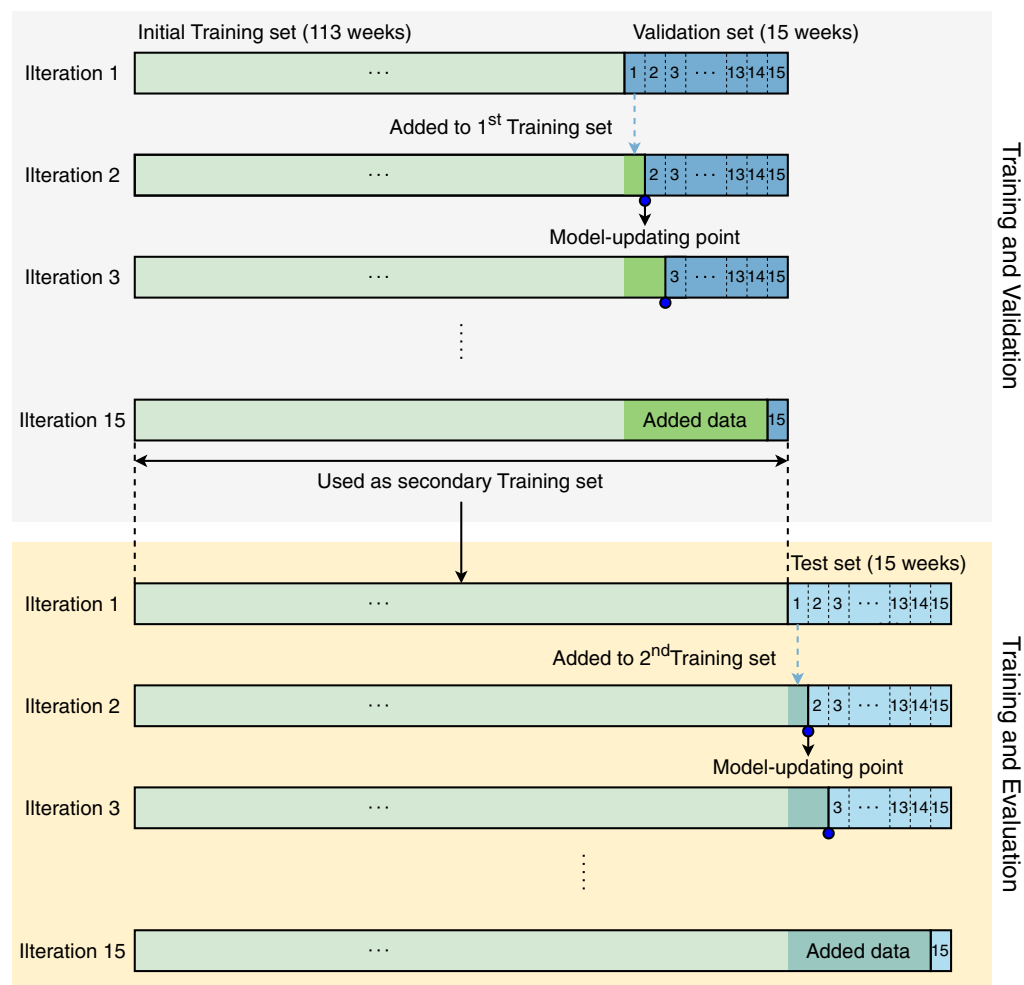


Figure 2 Modeling strategy in our study.

Full-size DOI: 10.7717/peerj-cs.2503/fig-2

last 15 weeks, the models in the *training and evaluation* phase were also updated 15 times. All validation and evaluation results were recorded for further analysis.

Training our model

Our model was optimized using the Adam optimizer. The time required to complete one training epoch varies from 200 to 250 s. In the *training and validation* phase, the first model (at iteration 1) was trained over 30 epochs with a learning rate of 0.001. The next updated models were trained with an additional 1-5 epochs. The training process was terminated when the validation loss reached the bottom and started to rise. In the *training and evaluation* phase, the first model was retrained with a secondary training set including all samples of training and validation data. After the first models were obtained, the next updated models were retrained with one additional epoch and evaluated by the data of the next week. The training process of the second phase was stopped after 15 iterations.

Training machine learning models

To perform comparative analysis, we trained four other models using conventional machine learning algorithms, including k -nearest neighbor (k -NN), linear regression (LR), random forest (RF), and eXtreme gradient boosting (XGB). Two deep learning models based on gated recurrent unit (GRU) (Elsworth & Güttel, 2020), and long short-term memory (LSTM) (Velarde et al., 2022) models were also implemented for compassion. A deep learning model based on TabTransformer (Huang et al., 2020) were trained to serve as a baseline to compare with our model. Each method leverages a distinct principle of learning from data. k -nearest neighbor works on the principle of proximity, predicting the label of a query point based on the most frequent label (classification) or average value (regression) of its k closest neighbors in the feature space (Nguyen, Tay & Chui, 2015). Linear regression works by fitting a linear equation to observed data, establishing a relationship between one dependent variable and one or more independent variables to predict outcomes. Random forest improves on the decision tree method by creating an ensemble of trees where each tree is trained on a subset of the data and features, thus reducing variance and avoiding overfitting (Nguyen-Vo et al., 2019). eXtreme gradient boosting enhances traditional gradient boosting by optimizing the algorithm's speed and efficiency, using a more regularized model formalization to control over-fitting, which makes it robust and highly accurate even on large and complex datasets (Pham et al., 2019). Gated recurrent unit and long short-term memory are two effective and commonly used recurrent neural networks for learning sequential data (Le et al., 2019).

Assessment metrics

To evaluate the performance of all models, we used mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2), and mean absolute percentage error (MAPE). The mathematical formula of these metrics are expressed as:

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

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{x})^2}{\sum_{i=1}^N (y_i - \bar{x})^2}, \quad (5)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (6)$$

where N is the total number of samples, \hat{y}_i is a predicted value, y_i is a ground truth value, and \bar{x} is a mean value over the total samples. These metrics are commonly used to compute the prediction error of regression models.

Table 2 Performance of our model compared to other machine learning models.

Model	MSE	RMSE	MAE	MAPE	R ²
<i>k</i> -nearest neighbor	213.71 ± 4.65	14.62 ± 1.54	8.68 ± 0.56	72.53 ± 2.35	0.55 ± 0.02
Linear regression	434.88 ± 10.12	20.85 ± 1.81	14.64 ± 0.87	166.71 ± 3.36	0.09 ± 0.01
Random forest	64.56 ± 2.11	8.03 ± 1.24	5.27 ± 0.54	60.99 ± 2.17	0.87 ± 0.03
eXtreme gradient boosting	239.44 ± 3.59	15.47 ± 1.67	9.61 ± 0.75	102.68 ± 2.95	0.5 ± 0.02
GRU-based model	37.84 ± 3.23	6.15 ± 1.14	3.76 ± 0.41	16.08 ± 1.25	0.92 ± 0.09
LSTM-based model	34.64 ± 4.68	5.89 ± 1.39	3.56 ± 0.48	30.26 ± 1.82	0.93 ± 0.08
TabTransformer	85.72 ± 3.74	9.26 ± 1.37	6.21 ± 0.59	79.89 ± 2.26	0.82 ± 0.09
Ours	25.76 ± 2.47	5.08 ± 1.23	3.12 ± 0.37	9.49 ± 1.11	0.95 ± 0.08

Computing platform

In our study, all modeling experiments were conducted on a personal computer (AMD Ryzen 7 5800X 8-Core Processor 3.80 GHz and 4×8 GB RAM). The deep learning and machine learning models were implemented using PyTorch (Paszke et al., 2019) version 2.0.0 (CUDA Toolkit 11.7) and scikit-learn library (Pedregosa et al., 2011) version 1.4.1 in Python 3.11.8 environment.

RESULTS AND DISCUSSION

Model robustness

Table 2 summarizes the comparative analysis between our model and other conventional machine learning models over 15 iterations to assess the stability of all implemented models. The results show that our model obtains smaller values of MSE, RMSE, MAE, and MAPE and higher R² values compared to other machine learning models. In terms of MAE, our model is the best-performing, followed by the random forest, eXtreme gradient boosting, *k*-nearest neighbor, and linear regression models. The MSE, RMSE, and R² values of all models also follow the same ranking order. These findings reveal that our proposed model works more effectively than other machine learning models often employed to address similar problems. Additionally, the adaptivity of the model in processing and learning tabular data is observed. Our results indicate that the variation of all recorded measures is very small, indicating highly stable performance regardless of sampling. The means of the MAE, MSE, and RMSE values are 3.12, 25.76, and 5.08, respectively, while the R² value is 0.9462. Figure 3 shows the predicted values plotted against actual observations, indicating that the model performs more effectively in predicting lower values, while its predictive performance tends to decline for higher values. This is due to the fact that the dataset has more samples with lower values.

Limitations

Despite the promising results, our work also has limitations that need to be addressed to improve the robustness and applicability of our model. Since the model performance was evaluated on a specific medium-sized dataset, the proposed method may not fully capture the diversity and complexity of real-world problems, limiting the applicability of our

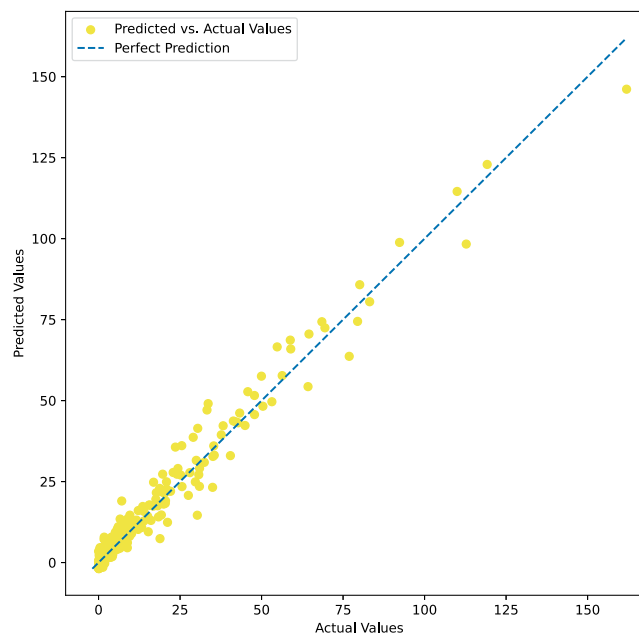


Figure 3 Predicted values plotted against actual observations.

Full-size DOI: 10.7717/peerj-cs.2503/fig-3

findings to similar topics. The comparative analysis is restricted to benchmarking our work against a small number of conventional machine learning models. In the future, more advanced models, including data-centric AI methods (Wang et al., 2024), can be implemented to further assess the model's performance.

CONCLUSION

In our study, we proposed a computational framework for enterprise sales forecasting using transformers, one of the most effective deep learning architectures. Our model was carefully developed to adapt to low-dimensional tabular data. Our findings suggest that our proposed method is highly robust and reproducible with low variation. Compared to other machine learning methods, our model showed better performance with lower MAE, MSE, and RMSE, and higher R^2 values of nearly 0.95. These results not only confirm the applicability of the model in this research but also its potential in similar studies using low-dimensional tabular data.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Social Science Foundation of China (Project No. 22BGL218), "On the Bubble Risk Identification and Prevention of Local Governments' Competition in Green Innovation under the Goal of Carbon Neutrality." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Social Science Foundation of China: 22BGL218.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yupeng Sun conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Tian Li conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The code and data used in the experiments are available in the [Supplemental File](#).

The data is originally from Kaggle competition “Walmart Recruiting-Store Sales Forecasting” available at: <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.2503#supplemental-information>.

REFERENCES

- Badaro G, Saeed M, Papotti P. 2023.** Transformers for tabular data representation: a survey of models and applications. *Transactions of the Association for Computational Linguistics* **11**(3):227–249 DOI [10.1162/tac1_a_00544](https://doi.org/10.1162/tac1_a_00544).
- Ban Y, Liu Y, Yin Z, Liu X, Liu M, Yin L, Li X, Zheng W. 2023.** Micro-directional propagation method based on user clustering. *Computing and Informatics* **42**(6):1445–1470 DOI [10.31577/cai_2023_6_1445](https://doi.org/10.31577/cai_2023_6_1445).
- Beheshti-Kashi S, Karimi HR, Thoben K-D, Lütjen M, Teucke M. 2014.** A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering* **3**(1):154–161 DOI [10.1080/21642583.2014.999389](https://doi.org/10.1080/21642583.2014.999389).
- Biswas B, Sanyal MK, Mukherjee T. 2023.** AI-based sales forecasting model for digital marketing. *International Journal of E-Business Research* **19**(1):1–14 DOI [10.4018/IJEER](https://doi.org/10.4018/IJEER).
- Boone T, Ganeshan R, Jain A, Sanders NR. 2019.** Forecasting sales in the supply chain: consumer analytics in the big data era. *International Journal of Forecasting* **35**(1):170–180 DOI [10.1016/j.ijforecast.2018.09.003](https://doi.org/10.1016/j.ijforecast.2018.09.003).
- Chen S, Ke S, Han S, Gupta S, Sivarajah U. 2024.** Which product description phrases affect sales forecasting? An explainable AI framework by integrating WaveNet neural network models with multiple regression. *Decision Support Systems* **176**(3):114065 DOI [10.1016/j.dss.2023.114065](https://doi.org/10.1016/j.dss.2023.114065).

- Chen I-F, Lu C-J. 2016. Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications* 28(9):2633–2647 DOI 10.1007/s00521-016-2215-x.
- Cheriyian S, Ibrahim S, Mohanan S, Treesa S. 2018. Intelligent sales prediction using machine learning techniques. In: *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. Piscataway: IEEE.
- Chu C-W, Zhang GP. 2003. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics* 86(3):217–231 DOI 10.1016/S0925-5273(03)00068-9.
- Das P, Chaudhury S. 2006. Prediction of retail sales of footwear using feedforward and recurrent neural networks. *Neural Computing and Applications* 16(4–5):491–502 DOI 10.1007/s00521-006-0077-3.
- Davis DF, Mentzer JT. 2007. Organizational factors in sales forecasting management. *International Journal of Forecasting* 23(3):475–495 DOI 10.1016/j.ijforecast.2007.02.005.
- Dharan BG. 1987. The effect of sales and collection disclosures on cash flow forecasting and income smoothing. *Contemporary Accounting Research* 3(2):445–459 DOI 10.1111/j.1911-3846.1987.tb00649.x.
- Elsworth S, Güttel S. 2020. Time series forecasting using LSTM networks: a symbolic approach. ArXiv preprint DOI 10.48550/arXiv.2003.05672v1.
- Fildes R, Beard C. 1992. Forecasting systems for production and inventory control. *International Journal of Operations & Production Management* 12(5):4–27 DOI 10.1108/01443579210011381.
- Forrest E, Hoanca B. 2015. *Artificial intelligence: marketing's game changer*. Hershey, PA: IGI Global, 45–64.
- Gao M, Xu W, Fu H, Wang M, Liang X. 2014. A novel forecasting method for large-scale sales prediction using extreme learning machine. In: *2014 Seventh International Joint Conference on Computational Sciences and Optimization*. Piscataway: IEEE.
- Gupta M, Kohli A. 2006. Enterprise resource planning systems and its implications for operations function. *Technovation* 26(5–6):687–696 DOI 10.1016/j.technovation.2004.10.005.
- He S. 2022. Auto sales forecasting model based on target customer satisfaction theory. In: *Proceedings of the 2022 International Conference on Mathematical Statistics and Economic Analysis (MSEA 2022)*. Amsterdam, Netherlands: Atlantis Press.
- Huang X, Khetan A, Cvitkovic M, Karnin Z. 2020. Tabtransformer: tabular data modeling using contextual embeddings. Available at <https://arxiv.org/abs/2012.06678>.
- Janczura J, Michalak A. 2020. Optimization of electric energy sales strategy based on probabilistic forecasts. *Energies* 13(5):1045 DOI 10.3390/en13051045.
- Kaggle. 2014. Walmart recruiting–store sales forecasting. Available at <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>.
- Kaneko Y, Yada K. 2016. A deep learning approach for the prediction of retail store sales. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. Piscataway: IEEE.
- Kohli S, Godwin GT, Urolagin S. 2020. *Sales prediction using linear and KNN regression*. Singapore: Springer Singapore, 321–329.
- Lawrence M, O'Connor M. 2000. Sales forecasting updates: how good are they in practice? *International Journal of Forecasting* 16(3):369–382 DOI 10.1016/S0169-2070(00)00059-5.
- Lawrence M, O'Connor M, Edmundson B. 2000. A field study of sales forecasting accuracy and processes. *European Journal of Operational Research* 122(1):151–160 DOI 10.1016/S0377-2217(99)00085-5.

- Le NQK, Nguyen QH, Chen X, Rahardja S, Nguyen BP. 2019. Classification of adaptor proteins using recurrent neural networks and PSSM profiles. *BMC Genomics* 20(S9):966 DOI 10.1186/s12864-019-6335-4.
- Leow EKW, Nguyen BP, Chua MCH. 2021. Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation. *Expert Systems with Applications* 179(2):115060 DOI 10.1016/j.eswa.2021.115060.
- Lyu F, Choi J. 2020. The forecasting sales volume and satisfaction of organic products through text mining on web customer reviews. *Sustainability* 12(11):4383 DOI 10.3390/su12114383.
- Mao Y. 2024. TabTranSELU: a transformer adaptation for solving tabular data. *Applied and Computational Engineering* 51(1):81–88 DOI 10.54254/2755-2721/51/20241174.
- Mccarthy TM, Davis DF, Golitic SL, Mentzer JT. 2006. The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. *Journal of Forecasting* 25(5):303–324 DOI 10.1002/(ISSN)1099-131X.
- Nguyen BP, Tay W-L, Chui C-K. 2015. Robust biometric recognition from palm depth images for gloved hands. *IEEE Transactions on Human-Machine Systems* 45(6):799–804 DOI 10.1109/THMS.2015.2453203.
- Nguyen-Vo T-H, Nguyen QH, Do TTT, Nguyen T-N, Rahardja S, Nguyen BP. 2019. iPseU-NCP: identifying RNA pseudouridine sites using random forest and NCP-encoded features. *BMC Genomics* 20(S10):1474 DOI 10.1186/s12864-019-6357-y.
- Ni Y, Fan F. 2011. A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications* 38(3):1529–1536 DOI 10.1016/j.eswa.2010.07.065.
- Osadchiy N, Gaur V, Seshadri S. 2013. Sales forecasting with financial indicators and experts' input. *Production and Operations Management* 22(5):1056–1076 DOI 10.1111/poms.12022.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. 2019. Pytorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, eds. *Advances in Neural Information Processing Systems*. Vol. 32. San Francisco, CA: Curran Associates, Inc.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Pham HN, Koay CY, Chakraborty T, Gupta S, Tan BL, Wu H, Vardhan A, Nguyen QH, Palaparthi NR, Nguyen BP, Chua MCH. 2019. Lesion segmentation and automated melanoma detection using deep convolutional neural networks and XGBoost. In: *2019 International Conference on System Science and Engineering (ICSSE)*. Piscataway: IEEE.
- Punam K, Pamula R, Jain PK. 2018. A two-level statistical model for big mart sales prediction. In: *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*. Piscataway: IEEE.
- Ramos P, Santos N, Rebelo R. 2015. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing* 34(17):151–163 DOI 10.1016/j.rcim.2014.12.015.
- Ribeiro A, Seruca I, Durão N. 2017. Improving organizational decision support: detection of outliers and sales prediction for a pharmaceutical distribution company. *Procedia Computer Science* 121:282–290 DOI 10.1016/j.procs.2017.11.039.

- Snyder R. 2002. Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research* **140**(3):684–699 DOI [10.1016/S0377-2217\(01\)00231-4](https://doi.org/10.1016/S0377-2217(01)00231-4).
- Snyder RD, Koehler AB, Ord J. 2002. Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting* **18**(1):5–18 DOI [10.1016/S0169-2070\(01\)00109-1](https://doi.org/10.1016/S0169-2070(01)00109-1).
- Sohrabpour V, Oghazi P, Toorajipour R, Nazarpour A. 2021. Export sales forecasting using artificial intelligence. *Technological Forecasting and Social Change* **163**:120480 DOI [10.1016/j.techfore.2020.120480](https://doi.org/10.1016/j.techfore.2020.120480).
- Sugiarto VC, Sarno R, Sunaryono D. 2016. Sales forecasting using holt-winters in enterprise resource planning at sales and distribution module. In: *2016 International Conference on Information & Communication Technology and Systems (ICTS)*. Piscataway: IEEE.
- Thomassey S, Zeng X. 2018. *Erratum to: artificial intelligence for fashion industry in the big data era*. Singapore: Springer Singapore.
- Tsoumakas G. 2018. A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review* **52**(1):441–447 DOI [10.1007/s10462-018-9637-z](https://doi.org/10.1007/s10462-018-9637-z).
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. 2017. Attention is all you need. ArXiv preprint DOI [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762).
- Velarde G, Brañez P, Bueno A, Heredia R, Lopez-Ledezma M. 2022. An open source and reproducible implementation of LSTM and GRU networks for time series forecasting. In: *The 8th International Conference on Time Series and Forecasting*. Basel, Switzerland: MDPI volume 323 of ITISE 2022, page 30.
- Wacker JG, Lummus RR. 2002. Sales forecasting for strategic resource planning. *International Journal of Operations & Production Management* **22**(9):1014–1031 DOI [10.1108/01443570210440519](https://doi.org/10.1108/01443570210440519).
- Wang AX, Chukova SS, Simpson CR, Nguyen BP. 2024. Challenges and opportunities of generative models on tabular data. *Applied Soft Computing* **166**(4):112223 DOI [10.1016/j.asoc.2024.112223](https://doi.org/10.1016/j.asoc.2024.112223).
- Weber FD, Schütte R. 2019. State-of-the-art and adoption of artificial intelligence in retailing. *Digital Policy, Regulation and Governance* **21**(3):264–279 DOI [10.1108/DPRG-09-2018-0050](https://doi.org/10.1108/DPRG-09-2018-0050).
- Xu A, Li Y, Donta PK. 2024. Marketing decision model and consumer behavior prediction with deep learning. *Journal of Organizational and End User Computing* **36**(1):1–25 DOI [10.4018/JOEUC](https://doi.org/10.4018/JOEUC).
- Yao B. 2023. Walmart sales prediction based on decision tree, random forest, and k-neighbors regressor. *Highlights in Business, Economics and Management* **5**:330–335 DOI [10.54097/hbem.v5i.5100](https://doi.org/10.54097/hbem.v5i.5100).
- Zhu C. 2023. An adaptive agent decision model based on deep reinforcement learning and autonomous learning. *Journal of Logistics, Informatics and Service Science* **10**(3):107–118 DOI [10.33168/JLISS.2023.0309](https://doi.org/10.33168/JLISS.2023.0309).
- Zhu D, Bahadur W, Ali M. 2023. The effect of spiritual leadership on proactive customer service performance: the roles of psychological empowerment and power distance. *Humanities and Social Sciences Communications* **10**(1):95 DOI [10.1057/s41599-023-02273-x](https://doi.org/10.1057/s41599-023-02273-x).