The American Journal of Management and Economics Innovations
ISSN 2693-0811 | Open Access

TYPE Original Research
PAGE NO. 59-65
DOI 10.37547/tajmei/Volume07Issue04-07

Transformers in Data-Driven Decision-Making: Applications for Forecasting Sales, Analyzing Demand, and Optimizing Pricing Strategies

Bulycheva Mariia

Senior Applied Scientist, Zalando Germany

OPEN ACCESS

SUBMITED 24 February 2025
ACCEPTED 27 March 2025
PUBLISHED 21 April 2025
VOLUME Vol.07 Issue 04 2025

CITATION

Bulycheva Mariia. (2025). Transformers in Data-Driven Decision-Making: Applications for Forecasting Sales, Analyzing Demand, and Optimizing Pricing Strategies. The American Journal of Management and Economics Innovations, 7(04), 59–65. https://doi.org/10.37547/tajmei/Volume07Issue04-07

Abstract: This article examines the methodological aspects of applying Transformer architectures for sales forecasting, demand analysis, and price optimization. The focus is on the development, adaptation, and integration of models in the context of processing large volumes of data and operating complex market mechanisms. The paper explores approaches to combining time series, identifying factor relationships, and improving the accuracy of analytical conclusions.

The methodology includes adapting basic Transformer architectures, such as Transformer with Multihead Attention Mechanism, to the specific characteristics of the data. The preparatory steps cover information aggregation, creation of temporal features, identification of categorical variables, and handling missing data. Historical datasets supplemented with external information sources are used for training. The sources include scientific articles by international authors published in open access, as well as materials available on the internet, allowing for a broad examination of the topic.

The results demonstrate the effectiveness of these architectures in forecasting tasks, identifying temporal dependencies, and improving business process quality. Examples of model implementation illustrate their successful use in commercial information systems. The conclusions emphasize the role of these approaches in decision-making automation and strategic planning.

The materials of the article are intended for professionals working in machine learning, data

analytics, and process management improvement.

Keywords: sales forecasting, demand modeling, price optimization, time series, machine learning, deep learning, Transformer architecture.

Introduction: The current realities of the global market require the development of accurate demand forecasting algorithms and the implementation of adaptive tools for price optimization. Managing these processes is fundamental to ensuring the competitiveness of companies. Machine learning methods, including deep learning networks, offer opportunities for analyzing relationships in large datasets that reflect both short-term fluctuations and long-term patterns.

Transformer models, initially designed for processing textual information, have proven to be versatile when working with temporal data and studying interactions between various variables. Their adaptation to time series analysis tasks enables the formation of consumer behavior forecasts, evaluation of marketing strategies' impact, and the development of approaches for demand forecasting and price management.

The use of machine learning technologies facilitates decision-making automation and improves forecasting accuracy in areas where the information is characterized by complex structure, seasonal changes, and multi-level dependencies. For the successful implementation of such solutions, it is important to consider data preparation aspects, model parameter optimization, and the analysis of the results obtained.

This article examines the application of Transformer architectures for sales forecasting, demand modeling, and price strategy development. Their effectiveness in integrating with business processes is analyzed.

MATERIALS AND METHODS

Modern approaches to the development of machine learning models for sales forecasting, demand analysis, and price determination are based on the use of deep neural networks, including Transformer with Multihead attention mechanism architecture and its derivatives. Integration with other methods is often employed, providing a variety of solutions. For the purpose of organizing the information, thematic areas were identified: sales forecasting, demand analysis, price modeling, and the integration of forecasts with optimization.

Sales forecasting is considered through the lens of transformer approaches. In the work of Cui E. et al. [4], a hybrid Transformer-BiGRU algorithm was proposed to account for temporal characteristics when processing sales data. The study by Xiang Y. et al. [7] presents a model that integrates temporal and frequency methods to achieve accuracy. Mu S. et al. [3] emphasized the flexibility of transformer computations used for sales analysis.

Demand analysis is explored by combining various methods. In the paper by Amellal I. et al. [1], an algorithm combining BERT, GRU, and probabilistic approaches for time series analysis is developed. Taparia V. et al. [11] describe the integration of regression models with machine learning algorithms, facilitating the processing of retail sales data. The work by Smirnov P. S. and Sudakov V. A. [12] presents an adaptive method used for predicting demand for new products.

Price modeling includes the use of flexible algorithms. A probabilistic Transformer for forecasting electricity prices is described in the work of Celeita Rodriguez D. F. [5]. Decision Transformer for real-time data analysis is presented in the paper by Zhang Z. and Wu M. [6]. Zhong B. [8] proposed a solution combining LSTM, ANN, and Transformer to integrate demand and price analysis.

The integration of forecasting with optimization is discussed in Zhang J. and Zhao J. [10], where sequential solutions for managing production, warehouse, and sales processes are considered. The approach highlights the importance of combining analytical models with management tasks. Li Q. and Yu M. [2] developed a sales forecasting model based on a modified Transformer architecture. The paper focuses on improving time sequence processing methods using technologies that optimize computational resources while improving accuracy. Zhou H. et al. [9] introduced the Informer algorithm designed for time series analysis. The study reveals a mechanism that reduces data density, which helps decrease computational costs without compromising accuracy. The new attention system architecture improves the algorithm's adaptation for tasks requiring the processing of large data volumes.

Transformer architectures are characterized by complexity in interpretation and computation. Issues related to processing irregular time series, adapting algorithms for new data, and improving model resilience to data changes remain relevant. The integration of analytical solutions with business processes requires further development aimed at practical application.

The work was written using an analytical methodology based on a systematic approach to data collection, study, and summarization.

RESULTS AND DISCUSSION

For time series forecasting, the Transformer architecture is adapted to the specific requirements of analysis. Modifications such as Temporal Fusion Transformer or Informer implement approaches that align with the data characteristics of this type.

The Temporal Fusion Transformer uses multidimensional attention, combines it with long-term memory networks, and provides event-level and feature-level dependency analysis. The model is characterized by interpretability.

The Transformer is focused on reducing computational complexity. The use of modified attention and probabilistic filtering ensures the processing of long sequences [4]. The process of building machine learning models with this architecture for sales forecasting, demand analysis, and price optimization is illustrated in Figure 1.

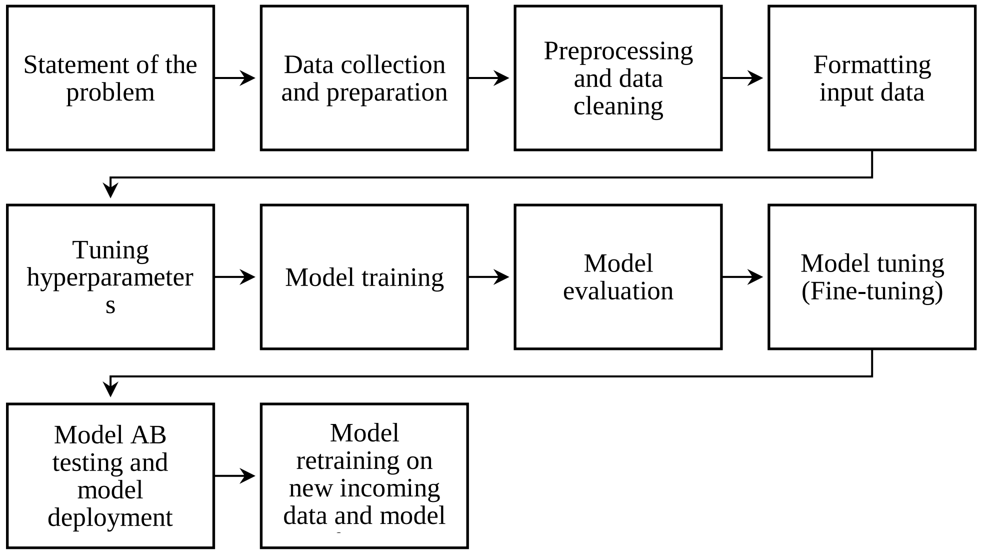

Figure 1. Stages of building machine learning models with Transformer architecture for sales forecasting, demand analysis, and price optimization [3,7].

Training takes into account interactions between parameters, including historical sales and demand, price and discounts changes, seasonal effects, characteristics of the products themselves, and the impact of marketing activities. To handle anomalies arising from unique events, contextual features are added. The models predict product sales behavior, identify trends in categories and regions. This approach is used to analyze price elasticity and evaluate relationships within the product matrix.

The configuration involves selecting standard Transformer parameters such as the number of attention heads, number of attention layers, model and feed forward dimensions. Forecast quality is assessed using metrics adapted to the analysis tasks: Mean Absolute Percentage Error, Root Mean Squared Error, and Weighted Mean Absolute Error.

The self-attention function identifies relationships between time intervals and allows for the analysis of patterns. Data integration, including macroeconomic indicators, pricing parameters, and competitor data, expands the capabilities of the models.

Price optimization involves determining the price of goods or services to increase profitability. The models process sales, demand, and competitor pricing data to identify patterns. This allows companies to set prices adapted to current market conditions and consider consumer behavior. Machine learning methods using transformers contribute to processing data on consumer preferences [1,5,12].

Below is an example of code using the PyTorch library to create a Transformer model:

```
import torch
import torch.nn as nn
import torch.optim as optim

class TransformerModel(nn.Module):
```

```
    def __init__(self, input_size, embed_size, num_heads, num_layers, output_size):
        super(TransformerModel, self).__init__()
        self.embedding = nn.Embedding(input_size, embed_size)
        self.transformer = nn.Transformer(embed_size, num_heads, num_layers, num_layers)
        self.fc_out = nn.Linear(embed_size, output_size)

    def forward(self, x):
        x = self.embedding(x)
        x = self.transformer(x, x)
        x = self.fc_out(x)
        return x

# Model parameters
input_size = 100  # Input size (e.g., number of products)
embed_size = 128  # Embedding size
num_heads = 8     # Number of attention heads
num_layers = 6    # Number of Transformer layers
output_size = 1   # Output size (forecasting demand or price)

model = TransformerModel(input_size, embed_size, num_heads, num_layers, output_size)

# Example training
criterion = nn.MSELoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

# Example input data
input_data = torch.randint(0, input_size, (10, 20))  # 10 samples, 20 time steps
output_data = torch.randn(10, 20, output_size)

# Training loop
model.train()
for epoch in range(10):
    optimizer.zero_grad()
    predictions = model(input_data)
    loss = criterion(predictions, output_data)
    loss.backward()
    optimizer.step()
    print(f'Epoch {epoch+1}: Loss = {loss.item()}')
```

For sales forecasting of a specific product, transformers are trained on data from previous periods. The input parameters include time series, information about holidays, seasonal fluctuations, and marketing activities. The model architecture includes an encoder that extracts hidden patterns from the input sequences and a decoder that transforms them into forecasts. The self-attention mechanism identifies dependencies between time points, which is necessary to account for data changes related to seasonality. The self-attention mechanism is based on calculating the importance of elements in the sequence relative to others [6,8]. Below is the formula that describes the attention mechanism:

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

Where:

- Q — query matrix;
- K — key matrix;
- K^T — transposed key matrix;
- V — value matrix;

- d_k — dimensionality of the keys.

The use of transformers in pricing and sales analysis tasks accounts for complex nonlinear dependencies in the data. Achieving results requires precise parameter tuning and consideration of the specifics of the input data. The integration of additional information sources, such as marketing campaign data and macroeconomic indicators, improves forecasts [10].

Next, in the context of this work, it is appropriate to consider the author's experience, during which a machine learning model was developed and subsequently implemented using the Transformer architecture. The goal of using this model is to forecast sales levels and analyze demand for specific products. Optimization of discount pricing strategies within the company, as well as the organization of inventory management, served as steps to control the growth rate, improve the company's top line and position among competitors in the market. The use of the Transformer architecture is due to its superiority over approaches based on recurrent neural networks and classic machine learning models, which is explained by the parallel processing of information, allowing the model training process to be accelerated. Additionally, a clear advantage of this type of network is its stability in performing tasks related to analyzing data input, considering changes in market conditions.

Regarding the development process, the system was first designed to transmit data, based on artificial intelligence algorithms. The data in this case included sales information, pricing policies, goods stored in warehouses, and external factors that could influence logistics processes. The automation in this case is focused on the analysis of data by the platform. The system then identifies existing dependencies in consumer behavior, which is necessary to understand changes in their preferences. This, in turn, allowed the optimization of inventory management, minimizing the unnecessary quantity of goods in warehouses, which directly reduced costs and helped avoid the risk of stock shortages.

As for the process of updating the platform, it is automated, ensuring timely adjustments to forecasts based on the uploaded data. This made it possible to derive optimal prices discounts based on forecasted demand levels in a timely and reliable manner. Below, Table 1 will describe the advantages and disadvantages of using machine learning models based on Transformer architecture for sales and demand forecasting.

Table 1. Advantages and disadvantages of Transformer architecture for sales and demand forecasting vs. other types of machine learning models (compiled by the author)

Aspect	Transformer Models	Classic ML Models	RNNs/LSTMs
Data Handling	Handle large-scale sequential and non-sequential data simultaneously. Efficient for multi-modal inputs (e.g., text, image, tabular).	Require feature engineering for temporal data. Limited ability to handle sequential data directly.	Effective for sequential data but struggle with large-scale or multi-modal datasets.
Temporal Dependencies	Capture long-term dependencies effectively via self-attention mechanisms.	Can miss long-term dependencies unless features are engineered manually.	Capture temporal dependencies but struggle with very long sequences due to vanishing gradients or limited memory.
Scalability	Scalable to large datasets due to parallelized processing.	Scalable but require significant preprocessing for complex datasets.	Limited scalability due to sequential nature of processing.
Performance	High accuracy when	Competitive for small	Strong performance on

	trained on sufficient data.	datasets or interpretable problems but may underperform on complex patterns.	small to medium sequential datasets but prone to overfitting on complex patterns.
Explainability	Difficult to interpret due to the "black-box" nature.	High interpretability with feature importance methods (e.g., SHAP, LIME).	Moderate explainability; harder to interpret than classic ML, but simpler than transformers.
Training Complexity	Computationally intensive; requires large-scale resources (GPUs/TPUs).	Relatively low complexity; can be trained on CPUs.	Moderate complexity; less computationally expensive than transformers but slower due to sequential training.
Data Requirements	Require large amounts of labeled data for effective training.	Perform well with smaller datasets if features are well-engineered.	Require significant labeled sequential data, but less than transformers.
Adaptability	Flexible; can handle different data types and tasks (e.g., multi-task learning).	Limited to predefined features and single-task learning.	Less flexible compared to transformers, primarily designed for sequence prediction tasks.
Handling Irregular Time Series	Can handle missing values and irregular intervals if preprocessed appropriately.	Struggle with irregular time intervals unless explicitly engineered.	Require padding or interpolation to handle missing or irregular data.
Memory and Hardware Needs	Require significant memory and high-end hardware for training and inference.	Efficient in terms of memory and hardware usage.	Moderate memory requirements but slower due to sequential nature.
Feature Engineering	Minimized need for feature engineering; learn representations directly from raw data.	High reliance on manual feature engineering for time-series and sales data.	Moderate; feature engineering is less intensive than classic ML but still required for specific use cases.
Generalization	Good generalization to unseen data if trained properly, but prone to overfitting with insufficient data.	Good generalization for well-engineered problems but limited adaptability.	Moderate generalization; better than classic ML for sequential data, but less robust to noise.

Thus, based on the above, it can be concluded Transformer models are the optimal choice for data analysis tasks involving large datasets, complex patterns, and long time series, as their ability to capture intricate dependencies and adapt to diverse data types ensures highly accurate forecasts—making them indispensable for driving data-informed decision-making in sophisticated and dynamic environments.

CONCLUSION

Based on the above, it should be noted that Transformer architecture is applied in tasks such as sales forecasting, demand estimation, and price formation. By analyzing time series, the models identify relationships within the data. Their application allows for accounting for changes and identifying key parameters. The popularity of their use is attributed to their ability to efficiently generate insights from large amounts of data that are subsequently applied in the decision-making process.

REFERENCES

Amellal I. et al. An integrated approach for modern supply chain management: utilizing advanced machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction //Decision Science Letters. – 2024. – Vol. 13(1). – pp. 237-248.

Li Q., Yu M. Achieving sales forecasting with higher accuracy and efficiency: A new model based on modified transformer //Journal of Theoretical and Applied Electronic Commerce Research. – 2023. – Vol. 18 (4). – pp. 1990-2006.

Mu S. et al. Transformative computing for products sales forecast based on SCIM //Applied Soft Computing. – 2021. – Vol. 109. – pp. 1- 13.

Cui E. et al. Fertilizer sales forecasting model based on transformer-BiGRU //Third International Conference on Machine Learning and Computer Application (ICMLCA 2022). – SPIE. - 2023. – Vol. 12636. – pp. 441-446.

Celeita Rodriguez D. F. An Intra-Day Electricity Price Forecasting Based on a Probabilistic Transformer Neural Network Architecture. – 2023. - №. 16. - pp. 1-24.

Zhang Z., Wu M. Real-time locational marginal price forecast: A decision transformer-based approach //2023 IEEE Power & Energy Society General Meeting (PESGM). – IEEE. - 2023. – pp. 1-5.

Xiang Y. et al. TSFRN: Integrated Time and Spatial-Frequency domain based on Triple-links Residual Network for Sales Forecasting //2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI). – IEEE. - 2023. – pp. 1012-1019.

Zhong B. Deep learning integration optimization of electric energy load forecasting and market price based on the ANN–LSTM–transformer method //Frontiers in Energy Research. – 2023. – Vol. 11. – pp. 1-14.

Zhou H. et al. Informer: Beyond efficient transformer for long sequence time-series forecasting //Proceedings of the AAAI conference on artificial intelligence. – 2021. – Vol. 35 (12). – pp. 11106-11115.

Zhang J., Zhao J. Prediction-Driven Sequential Optimization for Refined Oil Production-Sales-Stock Decision-Making //Energies. – 2022. – Vol. 15 (12). – pp. 4222.

Taparia V. et al. Improved Demand Forecasting of a Retail Store Using a Hybrid Machine Learning Model //Journal of Graphic Era University. – 2024. – pp. 15-36.

Smirnov P. S., Sudakov V. A. Forecasting new product demand using machine learning //Journal of Physics: Conference Series. – IOP Publishing, 2021. – Vol. 1925 (1). – pp. 1-8.