## **Transformer-Based Model for Predicting Customers' NPD in e-Commerce**

Subjects: Computer Science, Artificial Intelligence Contributor: Alexandru Grigoras, Florin Leon

Transformers offer advantages in capturing long-term dependencies within time series data through self-attention mechanisms. This adaptability to various time series patterns, including trends, seasonality, and irregularities, makes them a promising choice for next purchase day (NPD) prediction. The transformer model demonstrates improvements in prediction accuracy compared to the baselines.

e-commerce

transformer forecasting time series

next purchase day

## 1. Introduction

E-commerce is an important category of websites that help in connecting businesses with their customers. These platforms are constantly improving in order to meet the needs of customers. E-commerce represents a virtual store where customers can search, compare, and buy their favorite products from the comfort of their homes or on the go. Another important advantage is removing geographical barriers, helping businesses to have larger audiences and customers to get products that were not previously accessible without travelling to that location. By having a digital presence, companies can target customers more efficiently and understand their behavior when it comes to purchasing products. E-commerce has also created new buying trends, such as subscription models or personalized experiences [1].

Customer behavior is analyzed by e-commerce businesses to understand how their platform is used and what habits their customers have. User actions, purchases, trends, and preferences are used to gain valuable insights into their behavior. The analysis usually starts from the moment a customer lands on the website until their final purchase decision. Every click, search query, and scroll can provide important information. Using the insights generated from these, companies can create product recommendations and marketing campaigns that are helpful for both the customer and business. Another important aspect is determining the churn rate. This includes the abandoned shopping carts, lack of stock, reviews, and ratings, allowing companies to adjust their platform to fit their customers' needs. Studying consumer behavior is necessary for understanding the combination of requirements, expectations, and driving factors that influence customers [2].

There are multiple metrics that companies use to track their user behavior. These include conversion rate, average order value, customer lifetime value, customer acquisition costs, or others. These give businesses quantitative information about a variety of areas of their operations, such as sales, customer acquisition, and marketing performance. In this way, businesses can make data-driven decisions that have a solid foundation and remain flexible on small changes that occur depending on the market conditions and their customers' behavior <sup>[3]</sup>.

While the NPD metric can be used in multiple domains, such as "Software as a Service" (SaaS) products, financial services, hospitality, and supply chain management, the scope of this study is limited to e-commerce stores to better understand its characteristics and importance. Improving the method on an initial problem can be beneficial in generalizing a solution that can be used in multiple domains.

Predicting the NPD is addressed as a time series forecasting problem. The sequential purchases of each customer from e-commerce stores create a time series for the prediction task. The forecasting models usually require a significant amount of past data to estimate the future values. E-commerce platforms store these data internally, which can be used in predicting the next purchase day (NPD) <sup>[4]</sup>.

Knowing this metric can be used in multiple applications. For example, in targeted marketing, the NPD metric can be used to reach only the customers most likely to buy in the next period, optimizing the campaign budget. In revenue forecasting, the businesses can estimate future revenue using the number of customers that will purchase products as external information in the model. In customer lifetime value, businesses can better estimate and improve the future value of customers by determining their NPD, which can help in acquisition costs and determining the required marketing budgets.

## 2. Transformer-Based Model for Predicting Customers' NPD in e-Commerce

Predicting consumer behavior in e-commerce has attracted interest from businesses and academia.

In <sup>[5]</sup>, the NPD of customers from e-commerce websites is predicted using several methods, such as random forest (RF), autoregressive integrated moving average (ARIMA), convolutional neural network (CNN), and multilayer perceptron (MLP). The authors proposed a deep learning method based on long short-term memory (LSTM) networks. The model has two LSTM layers with sixteen neurons each. The authors test and compare the proposed model with RF, ARIMA, CNN, and MLP using a retail market dataset. While the model outperforms the alternative techniques in this study, it is shown that LSTMs have difficulty handling very long sequences due to the vanishing gradient problem, and scaling to larger datasets because of the lack of parallelization and sensitivity to hyperparameters. The studies <sup>[4][6]</sup> also determine the NPD for a customer on specific products using machine learning models, such as linear regression, XGBoost, ANNs, and RNNs. It uses the Cross Industry Standard Process for Data Mining (CRISP-DM) for building the solution to a feasible NPD predictor. A combination of methods are used, such as ANN with extreme gradient boosting (XGBoost), ANN with Recurrent Neural Network (RNN), and XGBoost with RNN. Overall, the ANN outperforms the individual models and the combinations of models with an error of less than 1–2 days and is selected for the final solution. ANNs are versatile for a large number of problems, but have limitations when handling sequential data, such as not taking into account positional information and lacking inherent memory, unlike RNNs.

Analyzing the papers that address the NPD problem, the usage of the same known methods for time series forecasting is observed. This leaves a gap for predicting the NPD with more efficient methods that handle sequential data. The following studies address different problems than NPD, but with more varied methods that show improvements in forecasting the specified datasets.

Thus, in a different context than NPD prediction, ref. <sup>[Z]</sup> compares ARIMA, LSTM and Prophet models for time series forecasting in the case of oil production datasets. ARIMA is found to be suitable for short-term predictions, while LSTM and Prophet excel in capturing unusual production changes. The Prophet model reveals potential seasonal effects. When extended to predict oil production across multiple wells in unconventional reservoirs, ARIMA and LSTM outperforms Prophet, suggesting the absence of seasonality in certain oil production curves.

Some hybrid approaches integrating deep learning models with statistical features extracted from the input data are described in <sup>[8]</sup>. They involve two stages: feature computation and XGBoost-based model fitting in the first stage, and training deep neural networks using original and feature-based samples in the second stage. Three deep learning methods, temporal convolutional networks (TCN), Multi-head Attention, and Multi-head-Attention-Res, are combined with XGBoost to create XGBoost-TCN, XGBoost-ATT, and XGBoost-ATT-Res models. The results on renewable energy consumption datasets reveal that the combined models outperform regular variants and seem to be superior to existing time series forecasting models, which demonstrates the effectiveness of integrating side information with input data for improved predictions.

The experiments performed in <sup>[9]</sup> also highlight the effectiveness of using hybrid approaches by combining two suboptimal models: ARIMA and XGBoost. This demonstrates that a fusion can produce forecasts comparable to or better than highly optimized models, in a significantly reduced computation time. Through extensive testing on over 4000 real-life time series, the hybrid method consistently delivers forecasts surpassing a heavily optimized ARIMA model while utilizing only around 25% of the computational resources. Furthermore, it outperformed the simpler ARIMA model with just a negligible increase in computational time. This underscores the importance of not only prioritizing forecasting accuracy but also considering advancements in speed and computational efficiency as time series forecasting gains prominence across diverse business domains and at increasingly granular levels.

The transformer architecture with its attention mechanism is used by researchers in different forecasting problems. By analyzing the production history of three wells, ref. <sup>[10]</sup> proposes a data-driven approach using an attention mechanism combined with a LSTM network (A-LSTM). This proves highly accurate and versatile; the combination significantly improves accuracy and accommodates noisy data well. It shows the stability, feasibility, and cost-efficiency of the A-LSTM compared to various other models, and it emphasizes A-LSTM as a powerful tool for production forecasting, promising valuable insights for engineers and operators in the oil and gas sector.

The research conducted in <sup>[11]</sup> addresses the need for accurate time series forecasting in diverse domains such as energy planning, epidemic prevention, and financial analysis. It identifies challenges tied to cumulative errors in autoregressive models for long-term forecasting and the complexity of temporal patterns. The proposed solution is a hierarchical transformer with a probabilistic decomposition representation, integrating the transformer with a

conditional generative model based on variational inference. This hierarchical approach effectively mitigates cumulative errors by imposing sequence-level constraints and enables the separation of temporal patterns for enhanced interpretability and prediction accuracy. The evaluation of multiple time series datasets shows the superior accuracy of the method compared to state-of-the-art approaches. Ablation experiments confirm the effectiveness and robustness of the probabilistic decomposition block, establishing this method as a reliable alternative for time series probabilistic forecasting.

Today's e-commerce stores are effective online platforms that give businesses access to a global consumer base, promote easy shopping experiences, and play an important part in modern commerce. Predicting the NPD of customers is an important metric that can help businesses understand and predict customer behavior, allowing for strategic planning and targeted marketing efforts to increase customer retention, enhance loyalty, and optimize sales and revenue in the long term.

While other studies on NPD prediction focus on known methods for time series forecasting, such as ARIMA, XGBoost and LSTM, a paper <sup>[12]</sup> proposes a new approach, based on the transformer model. The experiments were conducted on an online retail dataset, where a subset of customers were selected to determine their NPD. The proposed model was compared with the abovementioned techniques and shown to offer an improvement in RMSE of 6.15%. To further increase the model performance, the customers were grouped using the *k*-means clustering algorithm, demonstrating an additional improvement of 5.05% in RMSE, and a clear reduction in execution time. The advantages of using the transformer model are as follows: efficiently scaling on GPU because of parallelization, capturing long-term dependencies within time series data through self-attention mechanisms, and training the model on multiple time series by clustering to better capture the patterns in data. One of the limitations of the transformer model is that the self-attention mechanism is permutation-invariant and affects the order of the values returned, which impacts time series data more than in NLP applications.

The successful application of the transformer model in this context encourages further exploration and adoption of transformer-based solutions for a wide range of predictive analytics and forecasting tasks in the evolving ecommerce space. There are several paths to further expand the utilization of these models for predicting customers' NPD in the e-commerce domain. Addressing the challenge of imbalanced data present in customer purchase data is an important step. Integration of external features, such as marketing campaigns, seasonal trends, and customer demographics into the model, can improve the predictive accuracy by accounting for a broader array of influences on customer behavior. The proposed model should also be applied to multiple public benchmark time series to demonstrate its generalizability. Another direction of research is to explore the ensemble methods that combine the strengths of various predictive models, which could offer a robust predictive framework. Continuous model retraining is also important when applying a model in production, as it needs to remain effective in a dynamic landscape of e-commerce stores that are continuously changing. Enhancing interpretability and researching new methods of efficient retraining is also important for optimal results.

## References

- 1. What Is Ecommerce? A Comprehensive Guide (2023). Available online: https://www.shopify.com/blog/what-is-ecommerce (accessed on 1 September 2023).
- 2. Sharma, S.; Waoo, A. Customer Behavior Analysis in E-Commerce using Machine Learning Approach: A Survey. IJSRCSEIT 2023, 9, 163–170.
- 3. 15 Critical Ecommerce Metrics You Must Track (2023). Available online: https://www.shopify.com/blog/basic-ecommerce-metrics (accessed on 1 September 2023).
- Sharma, A.; Randhawa, P.; Alharbi, H.F. Statistical and Machine Learning Approaches to Predict the Next Purchase Date: A Review. In Proceedings of the 4th International Conference on Applied Automation and Industrial Diagnostics, Hail, Saudi Arabia, 29–31 March 2022.
- 5. Utku, A.; Akcayol, M. Deep Learning Based Prediction Model for the Next Purchase. AECE 2020, 20, 35–44.
- 6. Droomer, M. Predicting the Next Purchase Date for an Individual Customer using Machine Learning. Master's Thesis, Stellenbosch University, Stellenbosch, South Africa, 2020.
- 7. Ning, Y.; Kazemi, H.; Tahmasebi, P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. C&G 2021, 164, 105–126.
- 8. Abbasimehr, H.; Paki, R.; Bahrini, A. Novel XGBoost-Based Featurization Approach to Forecast Renewable Energy Consumption with Deep Learning Models. SUSCOM 2021, 38, 100–863.
- Lucaciu, R. Time Series Forecasting And Big Data: Can An Ensemble Of Weak Learners Decrease Costs While Maintaining Accuracy? In Proceedings of the 17th International Conference on Engineering of Modern Electric Systems, Oradea, Romania, 9–10 June 2022.
- 10. Indrajeet, K.; Bineet, K.T.; Anugrah, S. Attention-based LSTM network-assisted time series forecasting models for petroleum production. EAAI 2023, 123, 106–440.
- 11. Tong, J.; Xie, L.; Yang, W.; Zhang, K.; Zhao, J. Enhancing Time Series Forecasting: A Hierarchical Transformer with Probabilistic Decomposition Representation. IS 2023, 647, 119–410.
- 12. Grigoraș, A.; Leon, F. Transformer-Based Model for Predicting Customers' Next Purchase Day in e-Commerce. Computation 2023, 11, 210. https://doi.org/10.3390/computation11110210

Retrieved from https://encyclopedia.pub/entry/history/show/115410