DEMAND PREDICTION IN INDUSTRY 4.0 THROUGH A TRANSFORMER-BASED ARCHITECTURE

PREDICCIÓN DE LA DEMANDA EN LA INDUSTRIA 4.0 EMPLEANDO UNA ARQUITECTURA BASADA EN TRANSFORMERS

Sergio Joaquín González Herrera
Universidad Autónoma de Ciudad Juárez, México
al228199@alumnos.uacj.mx

José Mejía
Universidad Autónoma de Ciudad Juárez, México
jose.mejia@uacj.mx

Liliana Avelar Sosa
Universidad Autónoma de Ciudad Juárez, México
liliana.avelar@uacj.mx

Oliverio Cruz Mejía
Universidad Nacional Autónoma de México, México
oliverio.cruz.mejia@comunidad.unam.mx

Reception: 23/november/2023 **Acceptance**: 26/april/2024

Abstract

Given the changing circumstances faced by the industry today product demand is constantly influenced by several factors such as global economic conditions affected by wars, pandemics and recessions. For companies to cope with these challenges effectively, it is crucial to have a dependable demand prediction system that can be swiftly and efficiently communicated to their supply chain. However, this can be challenging for large consortiums with distributed supply chains spanning different countries. In this study a neural network architecture based on Transformers is proposed for demand prediction. This system could be integrated into a cloud service accessible to various locations within a company's supply chain, thus reducing information delays. By evaluating our approach with real product demand data and comparing it with other architectures the experiments prove that our model outperforms other methods.

Keywords: Demand prediction, Industry 4.0, Transformers.

Resumen

En la actualidad, la industria enfrenta desafíos debido a factores cambiantes como guerras, pandemias y recesiones que afectan la demanda de productos. Para enfrentar estos retos es crucial contar con un sistema confiable de predicción de demanda que se comunique eficientemente con la cadena de suministro. Sin embargo, esto puede ser un reto para grandes consorcios con cadenas de suministro distribuidas en diferentes países. En este estudio se propone una arquitectura de red neuronal basada en transformadores para predecir la demanda. Este sistema se integraría en un servicio en la nube accesible desde distintas ubicaciones dentro de la cadena de suministro, reduciendo los retrasos en la información. Los experimentos con datos reales demuestran que nuestro modelo supera a otras arquitecturas de predicción de demanda.

Palabras Clave: *Industria 4.0, Predicción de demanda, Transformadores.*

1. Introduction

The world's business landscape is characterized by rapid changes and disruptions in industrial strategies, affecting every aspect from product innovation and production processes to product delivery. The implementation of Industry 4.0 has proven to be particularly challenging for small and medium-sized manufacturing enterprises [Saptaningtyas, 2020]. Consequently, operations management faces daily challenges and numerous concerns that keep them awake at night, as they grapple with complex decisions to ensure the primary goals of corporate survival and growth [Behie, 2023].

In the face of various global challenges, such as the Economic Recession, Coronavirus Pandemic, Political Instability and Climate Change, businesses are compelled to seek agile solutions and make swift decisions [Kaushik, 2020], [Novoszel, 2022]. These changes have significantly affected demand prediction, which used to be stable and often focused on a single supplier. However, in the current industry landscape, businesses have multiple clients, located worldwide. As a result, exact demand prediction has become essential. Furthermore, it is crucial to send this information throughout all stages of a company's supply chain, which might also be globally distributed. Despite the challenges, demand forecasting is still a critical element in logistics and supply chain management.

Researchers have developed various approaches for demand prediction in the supply chain [Seyedan, 2020]. For instance, in one study [Kilimci, 2019], an intelligent demand forecasting system was created, relying on the analysis and interpretation of historical data. The model blended support vector machines with other models, adapted as a boosting ensemble for demand forecasting. In another work [Bandara, 2019], demand forecasts were proposed, leveraging sales correlations and relationships. The researchers compared two different Long-Short Term Memory (LSTM) learning schemes, incorporating both static and dynamic features.

In a different study [Vo, 2021], authors forecasted the demand for long-sleeved shirts using an Autoregressive Integrated Moving Average (ARIMA) model. Additionally, they used the forecast data to develop a Fuzzy economic production quantity inventory model. Moreover, in [Chandriah, 2021], a proposal involved using Recurrent Neural Networks (RNN)/LSTM with a modified Adam optimizer to predict the demand for spare parts. In this work, weight vectors were generated from data, and subsequently, the weights were optimized using the Modified-Adam algorithm.

The Transformer architecture is widely used in Natural Language Processing (NLP), a field in which it has proven to be highly effective [Koohfar, 2023]. Due to its effectiveness in working with sequential data, various researchers have extended the use of this architecture to other fields where datasets are structured as sequences and the expected results have a similar format.

Therefore, Transformers have appeared as a promising approach for demand forecasting. In one study [Zhang, 2023], a Transformer model was introduced to forecast intermittent demand, proving its superiority compared to other methods. In the context of tourism demand forecasting [Wu, 2023], a Temporal Fusion Transformer was trained using an adaptive differential evolution algorithm on tourism demand data. This Transformer, an attention-based deep learning model, effectively fused prediction with time-dynamic interpretable analysis, delivering robust results. Furthermore, researchers proposed a Transformer-guided probabilistic electricity demand forecasting framework in [Cao, 2022]. This framework learned both the global-local electricity demand dependencies and the complex correlations of external features associated with electricity demands. As a result, their proposed model could predict electricity demand distributions with a probabilistic state space model.

In the study [Wang, 2022], the researchers departed from traditional demand models that work well with linear time series. However, demand traffic models are rarely linear time series, so they created a model to forecast online ride-hailing demand traffic in 65 different areas, incorporating demand-related features such as weather conditions, local traffic conditions, PM 2.5 levels, and POI value. For this purpose, they employed an architecture that combines a Transformer and an LSTM to extract spatial and temporal features from a demand dataset. These features were used for bagging ensemble learning, followed by making predictions. At the end of the study, the results demonstrated that their proposal provides better forecasts in terms of MSE compared to those provided by ARIMA, ARIMAX, and RNN.[Wang, 2022] employed the Transformer architecture to predict strategic flight departure demand in various locations, aiming to support the performance of a mobile app. The model provides information about the expected flight departure demand to flight operators, offering a better overall view during specific times and reducing the number of delayed flights during peak airport activity periods.

Their study enhances the accuracy of current predictions. They utilized two different data sources: Aviation System Performance Metrics (ASPM) and System Wide Information Management (SWIM) to train forecasting models with Temporal Fusion Transformers (TFT) for five different airports.

This Transformer architecture was specifically designed to combine high-performance multi-scenario predictions with explanatory insights into temporal dynamics. Their proposed model achieved better results compared to Autoregressive models, 53%, and Linear Regression models, 31%.

[Koohfar, 2023] proposes a Transformer-based forecasting model to predict electric vehicle charging demand for periods of 7, 30, 60, and 90 days. This research is one of the first of its kind, considering the increasing use of electric vehicles in various parts of the world, which could lead to a significant shortfall in the availability of electrical energy. The study compares Deep Learning and Autoregressive methods against the Transformer model proposed by [Vaswani, 2017]. They used 1425 data points of charging demand and weather features in a city in Colorado, USA, from January 1, 2018, to December 28, 2021. The Transformer-based prediction model outperformed the autoregressive models in terms of RMSE and MSE metrics, with an RMSE of 0.055 vs 0.831 for short-term periods and 0.085 vs 0.920 for long-term periods.

In conclusion, the current industrial landscape demands agile strategies and decision-making due to the dynamic and disruptive nature of the world. Despite the challenges brought on by global factors, demand forecasting continues to be of paramount importance in effectively managing supply chains and ensuring business success. Researchers have explored various methods, including intelligent systems, LSTM approaches and Transformers with promising results, making significant strides in improving demand prediction accuracy and efficiency.

The current demand forecasting methods primarily focus on classical approaches like the ARIMA model or deep learning algorithms such as LSTMs or GRUs. In this research paper, a novel approach is proposed, using a Transformer model for time series prediction.

Originally designed for Natural Language Processing (NLP), the Transformer architecture has recently been explored for time series prediction, making it suitable for demand forecasting as well. In this work several modifications were made to an earlier Transformer model used in a different study [Zhang, 2023]. These modifications include introducing an input representation based on embeddings and incorporating a stack of dense networks with ReLU activation functions for the final forecast.

The key contributions of this research are twofold:

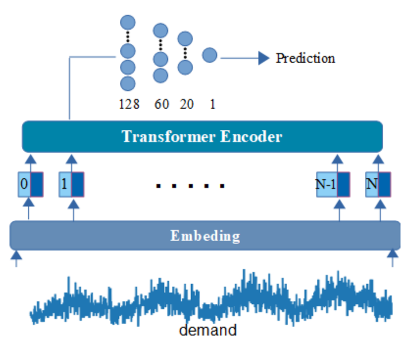
- A new deep learning Transformer-based architecture specifically designed for demand forecasting in the context of Industry 4.0.
- An empirical evaluation of the proposed framework using real-world retail sales data.

The next sections of the paper are organized as follows: In the Methods section the Transformer-based architecture is presented and described along with the evaluation metrics employed in the study. The Results section highlights the outcomes derived from the experiments conducted.

In the Discussions section, the obtained results are critically reviewed in the context of the proposed method. Finally, the Conclusions section presents the authors' last thoughts and conclusions about the study's findings.

2. Methods

Figure 1 displays a block diagram of the proposed network, which is based on a Transformer architecture [Vaswani, 2017], originally developed for natural language processing. It is important to note that in our proposed architecture, we exclusively employ the encoder portion of the Transformer. The decoder section is substituted with a sequence of dense networks responsible for generating the final prediction estimate.



Source: self made

Figure 1 Proposed Transformer architecture.

The original Transformer architecture comprises both an encoder and decoder, primarily designed to address sequence-to-sequence tasks while effectively managing long-range dependencies through the utilization of an attention mechanism. A key innovation in the Transformer architecture is its exclusive reliance on self-attention for computing representations of input and output, entirely avoiding the use of recurrent neural networks, which were prevalent in prior approaches. Self-attention, or intra-attention, is a specific type of attention mechanism that establishes connections between various positions within a single sequence to compute attention weights. To further enhance computational efficiency, they incorporated self-attention into a parallelized and efficient structure known as the multi-head self-attention mechanism [Vaswani, 2017]. This mechanism processes multiple sets of data concurrently and then combines the outputs from each set to generate the final output. This approach significantly accelerates computation speeds.

The proposed Transformer architecture employs historical demand information as input. Initially, this input undergoes a transformation using embeddings, and later, it is fed into the attention mechanism of the Transformer. Finally, the prediction is generated through a series of dense layers.

In this study, we exclusively employ the encoder architecture of the Transformer model, which leverages an attention mechanism to efficiently map an input sequence $(d_1, \ldots, d_n)$, into a sequence of continuous representations. The attention mechanism allows the Transformer to focus on important sections of the input data. This mechanism involves queries (Q), values (V), and keys (K). For demand forecasting, historical demand data is used, where given a time step n, d_n, our objective is to forecast the demand value at the next time step '$n + 1$' using 'k' steps of historical data. Therefore, the input sequence x is given as $(d_{n-k+1}, \ldots,\ d_{n-1},\ d_n)$.

The attention processing begins by dividing the input sequence into time steps, and then it determines which time steps to focus or pay attention to. For each time step a query is made and the keys stand for the rest of the time steps. The attention weights for a given query are obtained using compatibility functions between the query and each key. This results in specific weights assigned to each key in the statement. The values in our proposed scheme are derived from the time steps of the demand historical data, and these values are multiplied by the attention weights obtained from the attention processing. In our implementation we use word embeddings as representations of the input time steps.

For the compatibility function we use the scale dot-product attention mechanism. This mechanism calculates the dot product, QK^T, between a key and a query to determine the weight or attention assigned to a value, as shown in equation 1.

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

Unlike the approach in [Zhang, 2023], in the proposed architecture the output is obtained from dense layers with 120, 60, and 20 neurons, each using the Rectified Linear Unit (ReLU) activation function. The final output comes from a dense layer with a single neuron and ReLU activation function, assuming that the demand is positive. These dense layers are fed by the output from the attention mechanism. During training the entire network is optimized using the mean square loss function, which is commonly used for regression problems. By employing this architecture and training approach, we aim to improve demand forecasting accuracy and leverage the benefits of Transformer-based models in predicting time series data.

In the context of performance evaluation several metrics were employed to quantify prediction errors, as they are commonly utilized in this field of study. The analysis assumes a dataset with N observations. The metrics are outlined below.

R^2 Score (Coefficient of Determination): The R^2 score, equation (2), represents the percentage of variability in the dependent variable that can be accounted for by the independent variable [Temizhan, 2022]. It serves as a measure of how well the predicted values align with the actual values. A higher R^2 score indicates a better fit of the model to the data.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

Where y_i are data points, $\hat{y}_i$ is the prediction for data i and $\bar{y}$ is the mean of the data, the formula the summation runs over N, the number of observations, which is valid for equations 2 thru 6:

- Maximum Error ($MaxError$): The $MaxError$, equation 3 is a straightforward metric that measures the maximum difference between the predicted and actual values in the dataset. It provides insights into the largest prediction deviation, offering a valuable indicator of the model's performance.

$$MaxError = max_{\hat{y}_i}(|y_i - \hat{y}_i|) \tag{3}$$

- Mean Absolute Error (MAE): The MAE, equation (4), is a metric used to assess the average magnitude of errors between the observed data and the predicted values [Hodson, 2022]. It offers a more interpretable measure of prediction accuracy, representing the average absolute difference between the predicted and actual values.

$$MAE = \frac{1}{N}\sum_{i}|y_i - \hat{y}_i| \tag{4}$$

- Mean Squared Error (MSE): The MSE, equation (5), is a widely employed metric in various tasks, including time series prediction [Karunasingha, 2022]. It quantifies the average of the squared differences between the predicted and actual values, offering a comprehensive measure of prediction accuracy.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5}$$

- Median Absolute Deviation (MAD): MAD, equation 6, is part of robust statistics and serves to measure the variability of a univariate sample of quantitative data [Elamir, 2022]. It represents the $median$ value of the absolute differences between the predicted and actual values, providing insights into the model's performance with respect to potential outliers.

$$MAD = median(|y_i - \hat{y}_i|) \tag{6}$$

These metrics play a crucial role in evaluating the accuracy and effectiveness of prediction models, enabling researchers to objectively compare different forecasting approaches and identify areas for improvement. The utilization of such evaluation criteria is essential in the realm of demand forecasting within the context of Industry 4.0, where accurate predictions are paramount to effective supply chain management and business operations.

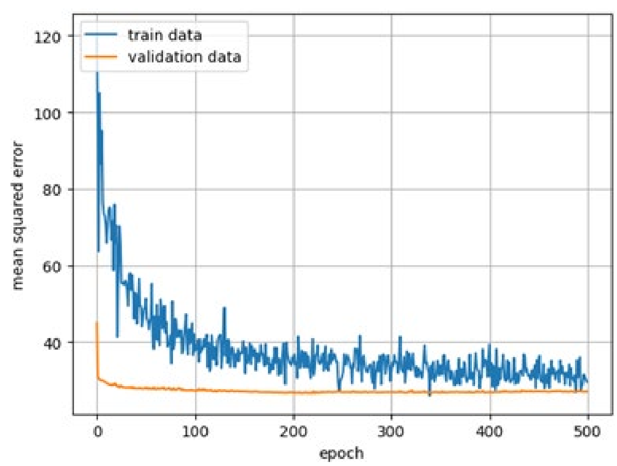
For the demand data, we extracted the first 1600 records from the "Store Item Demand Forecasting Challenge" dataset on Kaggle, which is free access. The information does not specify the type of the items sold; it only contains the sales dates, 1/1/2013 - 5/19/17, point of sale, item number, and the quantity sold on each sales date.

3. Results

The network's performance was evaluated using actual sales data for a specific product. The training process involved 80% of the data, and around 500 epochs were utilized, test data was 20%.

To assess the network's performance against alternative methods, a multi-layer perceptron (MLP) with five layers and the gradient boosting algorithm, commonly employed in time series prediction [Agapitos, 2017], [Körner, 2018], were also used for comparison.

In figure 2, the behavior of the proposed model during training is displayed using the Mean Squared Error (MSE) metric. The error curve gradually decreased as the number of epochs increased. It is noteworthy that the validation data was employed for adjustments in the initial epochs, where the validation data constituted 5% of the training data.

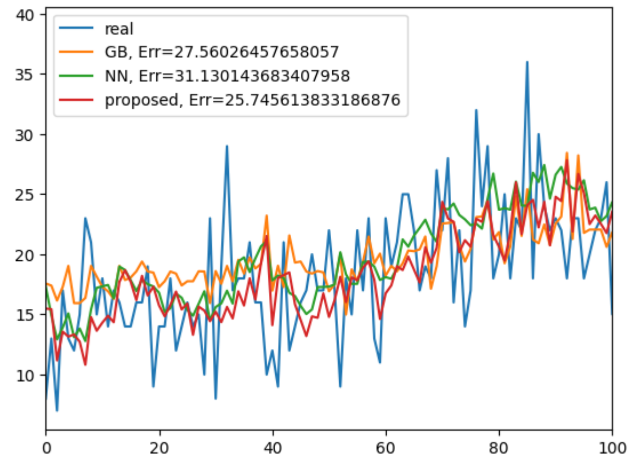


Source: self made

Figure 2 Mean Squared Error behavior during training.

Figure 3 illustrates the performance of the proposed architecture on the time series data, highlighting its MSE alongside the performance of the Gradient Boosting (GB) and Multi-Layer Perceptron (MLP) methods.

Table 1 provides a comprehensive overview of error metrics for various methods under evaluation. The results consistently prove the superior performance of the proposed method across all assessed metrics.



Source: self made

Figure 3 Different methods evaluated on test data.

Table 1 Error metrics.

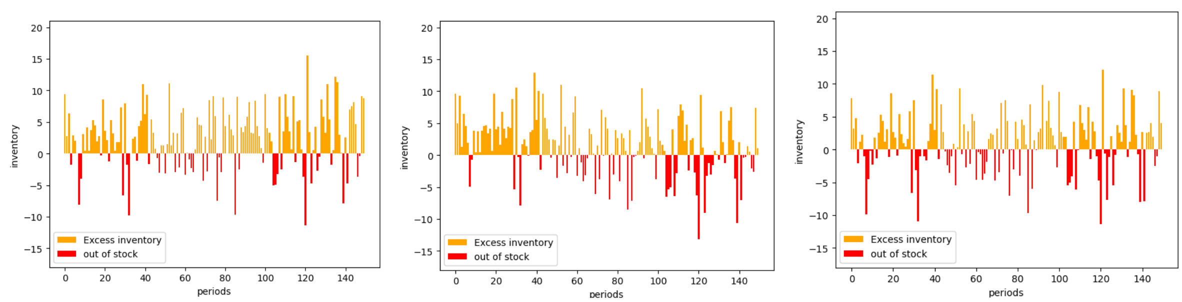
Method	R^2	$MaxError$	MAE	MSE	MAD
NN	0.2688	16.11	4.505	31.130	3.737
GB	0.3527	15.77	4.113	27.560	3.404
Proposed	0.3953	13.75	3.974	25.745	2.981

Source: self made

Specifically, the proposed approach reaches the highest value in the R^2 metric, signifying a strong correlation between predicted and actual values. Additionally, when scrutinizing metrics that gauge errors, the proposed method shows the lowest error values among all the compared methods. This observation shows the higher accuracy and precision of the proposed method in predicting the target variable compared to the alternatives.

The remarkable outcomes highlighted by the proposed approach underscore its efficacy in demand forecasting and highlight its potential to outperform other existing methodologies. Its ability to yield minimal errors in various error evaluation metrics reinforces its suitability for practical applications, where precise predictions play a crucial role in optimizing supply chain management and ensuring efficient resource allocation. The promising results obtained from this evaluation support the adoption of the proposed method as a workable solution to enhance demand forecasting performance in diverse industrial settings. Further investigations and real-world applications are warranted to validate and extend the findings presented here.

Furthermore, a case of inventory simulation is presented, wherein a subset of the time series was used, and Purchase Price and Sale Price were defined as 40 and 60, respectively. Figure 4 display the Missing Inventory and Excess Inventory using max(0, Actual Demand - Quantity Ordered) and min(0, Actual Demand - Quantity Ordered) respectively. The analysis suggests that the proposed method shows lower excess inventory, particularly during periods 0 to 40, resulting in reduced expenses for unsold products. On the other hand, the lack of inventory only leads to a loss of sales opportunities.

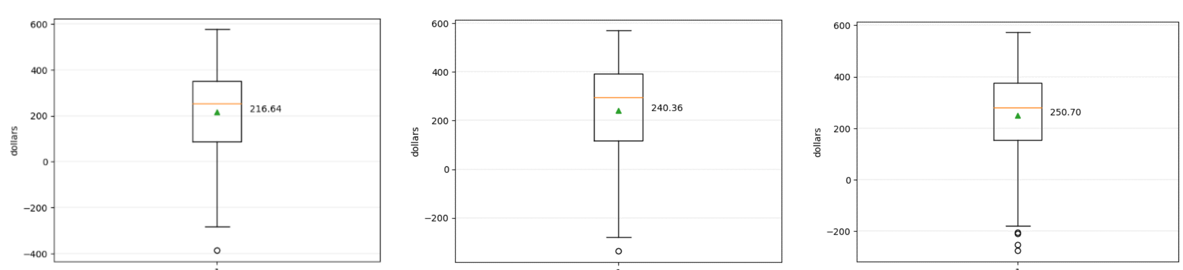


a) NN. *b) GB.* c) Proposed method.

Source: self made

Figure 4 Financial analysis, inventory cost.

In figure 5, the respective profit distributions for each methodology are visually represented using box plots. These plots employ horizontal lines to indicate the $median$ value within each box, while any data points falling significantly outside the box are identified as outliers, symbolized by small circles. It is evident that the proposed method achieves the highest profit compared to the other analyzed methods.



a) NN. *b) GB.* c) Proposed method.

Source: self made

Figure 5 Financial analysis, inventory cost.

The presented results substantiate the superiority of the proposed approach in demand forecasting, with its capability to minimize excess inventory and maximize profits, making it a favorable choice for inventory management in real-world applications.

4. Discussions

Based on the presented results, it is evident that the proposed approach outperforms other alternative methods, exhibiting superior quantitative performance as assessed by various metrics, including R^2, $MaxError$, MAE, MSE, and MAD. Figure 3 illustrates that while all methods capture the general trend of the time series, they do not precisely follow the fluctuations. This discrepancy might be attributed to the abrupt changes inherent in the time series data. Nevertheless, the proposed method excels in identifying and capturing the underlying trend, which accounts for its superior performance across the evaluated metrics when compared to the other methods.

Furthermore, in the supply chain experiment analysis, the proposed method also proves the best performance among the three methods. As depicted in the box plots shown in figure 5, the proposed method yields higher profits and shows lower variance in expected profit. This significant improvement is attributed to the superior prediction capabilities of the proposed method, enabling more accurate forecasting of demand for each period.

The observed outcomes validate the effectiveness of the proposed scheme in demand forecasting and its potential applicability in real-world scenarios. The ability of the proposed method to better capture the underlying trends of the time series data contributes to its advantages over alternative approaches, culminating in improved supply chain management and increased profitability. The findings presented here provide compelling evidence in support of the proposed method's viability and superiority, reinforcing its importance in advancing demand forecasting methodologies and optimizing industrial operations. To further solidify the conclusions drawn, more experiments and validations in diverse contexts are recommended to consolidate the promising implications of the proposed approach.

Considering the results obtained, employing a modified Transformer for demand prediction emerges as a promising choice when implementing a demand predictor in industrial settings. This is in line with other studies that have also utilized Transformers for forecasting several types of demand, such as Electric Vehicle Charging Demand in [Koohfar, 2023], Energy Consumption Demand in [Nazir, 2023], and Tourism Demand Forecasting in [Yi, 2021]. Transformers, owing to their incorporation of an attention mechanism, possess the unique ability to focus on crucial elements within input data. In our specific architecture, this feature enables a series of neural network layers to prioritize the most relevant elements, thus enhancing its predictions of future events. It is worth noting the that a simple projection function or dot product, in this case QK^T, see equation (1), which effectively establishes similarities between tensors, facilitating the assignment of attention weights used later in the dense block to make predictions. Note also that the proposed architecture deviates from the traditional Transformer design in that it eliminates the need for a decoding section. In the case of our system, the output from the Transformer's attention layer suffices for the dense network block to generate accurate predictions. This streamlined approach not only enhances computational efficiency but also makes it more feasible to implement in embedded devices like a Raspberry Pi, given its reduced computational demands.

5. Conclusions

This research introduces a novel approach for product demand prediction, using a new deep learning architecture based on Transformers. The proposed architecture incorporates a sequence of dense layers with ReLU activation for generating output predictions. Extensive experimentation was conducted to assess the performance of this architecture.

In the first experiment, the predictive capability of the proposed architecture was compared against alternative methods using several standard metrics commonly employed for evaluating time series predictions. The results revealed that the proposed architecture achieved the smallest error, represented by an MSE value of 25.745, while the closest competitor obtained an MSE of 27.56. This outcome signifies that the proposed network consistently delivers better predictions than the alternative predictors, proving its efficacy in demand forecasting.

Furthermore, a second experiment involved simulating a simple inventory system. In this scenario, the proposed network once again exhibited the best performance, reinforcing the connection between the network's accurate predictions and improved profit outcomes in inventory management. This finding is particularly valuable for manufacturing companies, offering enhanced production planning capabilities and other beneficial possibilities. As future work, we intend to further refine and adapt specific aspects of the Transformer service mechanism to better suit the time series generated by demand. One key focus is to enhance the distribution allocation process to align more effectively with the characteristics of demand time series.

6. Bibliography and References

[1] Agapitos, A., Brabazon, A., & O'Neill, M. Regularised gradient boosting for financial time-series modelling. Computational Management Science. Vol. 14, p. 367-91, 2017.

[2] Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26, p. 462-474, 2019.

[3] Behie, SW., Pasman, HJ., Khan, FI., Shell, K., Alarfaj, A., El-Kady, AH., & Hernandez, M. Leadership 4.0: The changing landscape of industry management in the smart digital era. Process Safety and Environmental Protection. No. 172, p. 317-328, 2023.

[4] Cao, Y., Dang, Z., Wu, F., Xu, X., & Zhou, F. Probabilistic Electricity Demand Forecasting with Transformer-Guided State Space Model. In 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). p. 964-969 IEEE, 2022.

[5] Chandriah, KK., & Naraganahalli, RV. RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand

forecasting. Multimedia Tools and Applications. No. 17, Vol. 80, p. 26145-26159, 2021.

[6] Elamir, EA. On Uses of Mean Absolute Deviation: Shape Exploring and Distribution Function Estimation. arXiv preprint arXiv:2206.09196. 2022.

[7] Hodson, TO. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions. p. 1-10, 2022.

[8] Kaushik, M., & Guleria, N. The impact of pandemic COVID-19 in workplace. European Journal of Business and Management. No. 15, Vol. 12, p. 1-10, 2020.

[9] Karunasingha, DS. Root mean square error or mean absolute error? Use their ratio as well. Information Sciences. Vol. 585, p. 609-29, 2022.

[10] Kilimci, ZH., Akyuz, AO., Uysal, M., Akyokus, S., Uysal, MO., Atak Bulbul, B., & Ekmis, MA. An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain. Complexity. Vol. 2019, Article ID 9067367, 15 pages, 2019.

[11] Koohfar, S., Woldemariam, W., & Kumar, A. Prediction of Electric Vehicles Charging Demand: A Transformer-Based Deep Learning Approach. Sustainability. No. 3, Vol. 15, p. 2105, 2023.

[12] Körner, P., Kronenberg, R., Genzel, S., & Bernhofer, C. Introducing Gradient Boosting as a universal gap filling tool for meteorological time series. Meteorol. Z. No. 5, Vol. 27, p. 369-376, 2018.

[13] Nazir, A., Shaikh, A. K., Shah, A. S., & Khalil, A. Forecasting energy consumption demand of customers in smart grid using Temporal Fusion Transformer (TFT). Results in Engineering, Vol. 17, 2023.

[14] Novoszel, L., & Wakolbinger, T. Meta-analysis of supply chain disruption research. Operations research forum (2022). Cham: Springer International Publishing. No. 1, Vol. 3, p. 10, 2022.

[15] Saptaningtyas, WW., & Rahayu, DK. A proposed model for food manufacturing in smes: Facing industry 5.0. Proceedings of the 5th NA International Conference on Industrial Engineering and Operations

Management Detroit, Michigan, USA, August 10 - 14, 2020. p. 1653-1661, 2020.

[16] Seyedan, M., & Mafakheri, F. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities, Journal of Big Data. No. 1, Vol. 7, p. 1-22, 2020.

[17] Temizhan, E., Mirtagioglu, H., & Mendes, M. Which correlation coefficient should be used for investigating relations between quantitative variables. American Academic Scientific Research Journal for Engineering, Technology, and Sciences. Vol. 85, p. 265-277, 2022.

[18] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., Kaiser, Ł., & Polosukhin, I. Attention is all you need. Advances in Neural Information Processing Systems 30 (NIPS 2017). Vol. 30, 2017.

[19] Vo, TT., Le, PH., Nguyen, NT., Nguyen, TL., & Do, NH. Demand Forecasting and Inventory Prediction for Apparel Product using the ARIMA and Fuzzy EPQ Model. Journal of Engineering Science & Technology Review. No. 2, Vol. 14, p. 80-89, 2021.

[20] Wang, L., Mykityshyn, A., Johnson, C., & Cheng, J. Flight demand forecasting with transformers. In AIAA AVIATION Forum 2022. p. 3708, 2022.

[21] Wang, T., Li, Y., Chang, W., & Zhou, S. A bagging ensemble learning traffic demand prediction model based on improved LSTM and transformer. Third International Conference on Computer Science and Communication Technology (ICCSCT 2022). Vol. 12506, p. 469–477, 2022.

[22] Wu, B., Wang, L., & Zeng, YR. Interpretable tourism demand forecasting with temporal fusion transformers amid COVID-19. Applied Intelligence. No. 11, Vol. 53, p. 14493-14514, 2023.

[23] Yi, S., Chen, X., & Tang, C. Tsformer: Time series transformer for tourism demand forecasting. arXiv preprint arXiv:2107.10977, 2021.

[24] Zhang, GP., Xia, Y., & Xie, M. Intermittent demand forecasting with transformer neural networks. Annals of Operations Research. p. 1-22, 2023.